DOCUMENT RESUME

ED 199 630 CS 005 900

AUTHOR	Miller, Wallace D., Ed.; McNinch, George H., Ed.
TITLE	Reflections and Investigations on Reading: Twenty-fifth Yearbook of the National Reading Conference.
INSTITUTION	National Reading Conference, Inc.
PUB DATE	76
NOTE	337p.; Proceedings of the National Reading Conference (25th, St. Petersburg, FL, December 4-6, 1975).
EDRS PRICE	MF01/PC14 Plus Postage.
DESCRIPTORS	Adult Education; Content Area Reading; Elementary Secondary Education; English (Second Language); Higher Education; Language Acquisition; *Language Skills; Literature Reviews; Minority Groups; Reading Attitudes; *Reading Comprehension; *Reading Instruction; Reading Materials; *Reading Research; Reading Skills; Reading Teachers; Study Skills; *Teacher Education; Teacher Effectiveness; Word Recognition
IDENTIFIERS	*National Reading Conference (Organization)

ABSTRACT

Focusing on the broad topic of reading, the articles in this volume are arranged into the following categories: literature reviews, college and adult education, teacher education, reading research, language, and comprehension. The literature reviews cover the areas of the underlining study aid, learning from text, sex roles and reading instruction, syntax and reading comprehension, reading and the aging process, and grammatical usage and reading. Among the topics dealt with in the remaining sections are: (1) prediction of academic success of special groups of university freshman, (2) integrating reading and study skills into college biology, (3) transferring reading and study skills to the content area, (4) the effects of testwiseness training on reading scores on test anxiety, (5) the relationship between performance and interest in adult education, (6) functional literacy of teacher trainees, (7) relationships in teacher proficiency, (8) the reading of secondary school American Indian students, (9) self-concept in reading instruction programs, (10) conceptual difficulty of reading material, (11) the influence of minority groups in the selection of basal reading programs, (12) the child's conception of a spoken word, (13) recognition of word boundaries by adult second-language subjects, (14) identifying sentence meaning skills, (15) a model for the comprehension of poetry, and (16) the effect of reader generated questions on learning from prose. (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Reflections and Investigations on Reading

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy

ED199630

TWENTY-FIFTH YEARBOOK OF THE NATIONAL READING CONFERENCE

1976

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
National Reading Conference, Inc.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Edited by
Wallace D. Miller
and
George H. McNinch
University of Southern Mississippi

CS05900

Published by
The National Reading Conference, Inc.
Clemson, South Carolina 29631

The ideas expressed in the papers published
in the Yearbook are those of the authors and
do not necessarily represent the viewpoints
of the National Reading Conference

Copyright 1976
The National Reading Conference, Inc.
All rights reserved. This book or parts thereof
may not be reproduced in any form without
permission of the publishers.

Distributed by
THE NATIONAL READING CONFERENCE
Godfrey Hall
Clemson University
Clemson, South Carolina 29631

PRINTED IN THE UNITED STATES OF AMERICA
BY
UNIVERSITY OF SOUTHERN MISSISSIPPI PRESS
Hattiesburg, Mississippi
$10.00

OFFICERS AND BOARD OF DIRECTORS
OF THE
NATIONAL READING CONFERENCE, INC.

Edward Fry, *President*
Rutgers University

J. Jaap Tuinman, *Vice-President*
University of Indiana

Gordon Gray, *Treasurer*
Clemson University

Voncile Mallory, *Secretary*
Florida Atlantic

Earl Rankin, *Past-President*
University of Kentucky

Martha Maxwell, *Member*
University of California, Berkeley

Wayne Otto, *Member*
Wisconsin University

Ron Carver, *Member*
University of Missouri

Richard Robinson, *Convention Manager*
University of Missouri

J. Jaap Tuinman, *Member &*
Journal of Reading Behavior Editor
University of Indiana

Wallace Miller and George McNinch, *Yearbook Editors*
University of Southern Mississippi

EDITORIAL ADVISORY COMMITTEE
25TH NRC YEARBOOK

Bill Blanton
Appalachian State University

Emery P. Bliesmer
Pennsylvania State University

Robert Chester
University of British Columbia

James Dinnan
University of Georgia

Ken L. Dulin
University of Wisconsin

Thomas H. Estes
University of Virginia

Martha Evans
University of Maryland

Thomas P. Fitzgerald
State Education Department of New York

Frank Greene
McGill University

Ann Marie Haase
University of Massachusetts

A.J. Kingston
University of Georgia

David Knight
Oglethorpe College

Karl Koenke
University of Illinois

Phil L. Nacke
University of Kentucky

Robert A. Palmatier
University of Georgia

Mark G. Richmond
University of Southern Mississippi

Richard Robinson
University of Missouri

Harry Singer
University of California

David Wark
University of Minnesota

John A. Wood
Kutztown State College

The Oscar S. Causey Award for Outstanding Achievement in College-Adult Reading

REMARKS OF HAZEL HORN CARROLL IN PRESENTING THE OSCAR S. CAUSEY AWARD TO A.J. LOWE ON BEHALF OF NATIONAL READING CONFERENCE

The recipient of the 1975 Oscar S. Causey Award is a unique combination of dependable practicality and scholarship. His efficiency is so outstanding that others have had to stay alert so as not to be over-dependent on him. And he is a "difficult act" to follow.

By now, all who have attended several NRC conventions know that I am referring to Al Lowe who was our Convention Manager and Registrar from 1963 to 1973.

While Al seemingly was effortlessly performing these duties so well he was pursuing diligently other facets of his professional life.

After this native of Massachusetts graduated from Silton School in New Hampshire, he received his B.A. degree in Psychology and Sociology and his M.A. degree in Reading at Mississippi Southern College. And he earned, also, a Doctor of Education degree in Psychology of Foundations and Services and Psychology of Reading at the University of Virginia.

He has taught Education and Psychology at Mississippi Southern and the University of Virginia. At Loyola University of New Orleans he had professional duties in Education and was Director of the Reading

Center. Now he is Associate Professor of Secondary Reading and Reading Research in the College of Education of the University of South Florida.

Al Lowe's over fifty publications have appeared in the AERA Research Journal, Research in Teaching English, University of Virginia Education Review, Reading Improvement, Reading Teacher, Journal of Reading, Perceptual and Motor Skills, College Student Journal, Psychology Reports, Journal of English Language Training, Florida Reading Quarterly, Florida Journal of Educational Research and the Yearbooks of the National Reading Conference. Also, he contributed a chapter in NRC's Lawrence Hafner's book *Reading in Middle and Secondary Schools.*

His research includes interesting studies such as: Surveys of College Reading Improvement Programs, Evaluations of College Reading Improvement Programs, Reviews of Literature in College-Adult Reading, Spelling Lists and Scales, Various Institutional Studies, Materials and Needs of Culturally Deprived Readers, Legibility, Critical Reading, Reading Tests, Vocabulary and Themes.

Many organizations have used A.J. Lowe as a Consultant. They include: Headstart, Upward Bound, AIM, Score, Learning Center, Vocational Rehabilitation, Prison, Model Cities, TESOL and others. His consulting has been for all school levels in addition to adult reading.

More research and articles are bound to come from Al Lowe because he is only 44 years old and he likes to communicate to his fellow educators as well as to lay people which he has done in newspaper articles, television shows and speeches.

Formerly he worked as a radio and television announcer and engineer. And he is a combat army veteran of the Korean War.

His versatility, his energy, his efficiency and his many interests have guided the numerous contributions to NRC that deserve to be recognized and honored. To A. J. Lowe, then, our thanks and appreciation symbolized by this Oscar S. Causey Award of 1975.

TABLE OF CONTENTS

PRESIDENTIAL ADDRESS

Who Are NRC? 1
Edward Fry

REVIEWS

A Critical Review of Research and Expert Opinion on the Underlining Study Aid 8
William G. Browning

Learning From Text: A Review of Theories, Stategies, and Research at the High School Level 22
Harry Singer and Alan Rhodes

Sex Roles and Reading Instruction: A Critique of Related Research 52
Patricia H. Duncan

Syntax and Reading Comprehension Do They Have a Future Together??? 63
Patricia M. Cunningham

The Aging Process and Its Relationship to Reading: A Review of Literature from Gerontology with Implications for Future Research 70
Richard D. Robinson and Gerald Maring

Grammatical Usage and Reading: A Review of the Literature . . 79
Mary F. Franke and Earl F. Rankin

COLLEGE AND ADULT

Prediction of Academic Success of Special Groups of University Freshmen 88
Joseph A. Muia and Emery P. Bliesmer

Prediction of College Success: A Test of Standard Measures for Traditional and Open Admissions Students 96
Josephine S. Goldsmith and Mark J. Nicolich

Integrating Reading and Study Skills Into College Biology . . . 102
Barbara Tomlinson and Michael D. Tomlinson

Transferring Reading and Study Skills to the Content Area . . 104
Teryle Carpenter and Yvonne Jones

Measurement of College Students' Self Concept: A Pilot Study 108
Ann Marie Bernazza Haase, Mervin D. Lynch, and Shirley DeShields

The Effects of Testwiseness Training on Reading Scores and on Test Anxiety 113
Gloria Batkin Kahn

College Students' Perceptions of Cheating and Non-Cheating Behaviors During Test-Taking 114
Ken L. Dulin and M. Jane Greenewald

A Comparison of Verbal Performance on Foreign and Native Born Graduate Students 115
Elaine Crable

Evaluation of Reading Improvement Courses: The Carver-Darby Chunked Reading Test Compared with Tests Prepared for Use with the Course Materials 119
Michael Bradley

The Reading Interests, Reading Attitudes, and Reading Backgrounds of Private-College Incoming Freshmen . . . 120
George J. Cretilli, Ken L. Dulin, and Robert Chester

A Systematic Approach to Predicting Reading Performance at the Secondary Level 121
Charles W. Peters and Nathaniel A. Peters

Cognitive and Personality Predictors of Listening Ability in College Students 129
Jane Glotfelty Rhoads, Patricia Ellis, and E.B. Sweney

The Need for Study Skills as Perceived by Community College Students 131
Dorinda Ann McClellan and Leslie McClellan

Relationship Between Performance and Interest in Adult Education 133
Thomas P. Fitzgerald

TEACHER EDUCATION

An Instrument Designed to Identify Specific Strengths and Weaknesses in Reading Instruction 135
Bob W. Jerrolds, Sidney E. Beton, and Judy Bazemore

Ascertaining Knowledge of Reading with the Artley-Hardin Inventory 141
Karl Koenke

Functional Literacy of Teacher Trainees 147
Emilie P. Sullivan

The Needs Assessment for Reading Teachers: A Comparison of the Diagnostic and Prescriptive Dimensions in Reading . . 152
Martha H. Dillner and Carl N. Shaw

A Study of Relationships in Teacher Proficiency 154
William R. Powell

How Necessary Are Certain Skills for Reading Competence? . . 161
Marie Carducci-Bolchazy

Reading Attitude and Achievement of Teacher Education Students 163
Lawrence Erickson and Thomas Hatcher

Symposium: Reading Problems and Habits of American Indian Youth 164
Emery P. Bliesmer

Secondary Indian Youth and Their Reading 166
Billie Sue Allen

Implications for the Teaching of Reading to American Indian Students in the Elementary Grades 170
Sandra J. Fox

Observations on the American Indian College Student, with Implications for Reading 175
John W. Tippeconnic

Increasing Phonics Proficiency of Pre-Service Teachers Through Modularized Instruction 180
Henry D. Olsen

RESEARCH

A Critical Evaluation of Several Methods for Measuring Gain in Reading 181
Earl F. Rankin and Lowell Eberwein

Reading and Other Language Skills of Residents in a State School and Hospital 190
Emery P. Bliesmer, Vincent S. Orlando, and Joseph A. Muia

An Examination of Student Coping Style, Teacher Control and Student Achievement in Reading 196
Homer Coker and Jeffrey Lorentz

Self Concept in Reading Instruction Programs 198
Mervin D. Lynch and Ann Marie Haase

A Longitudinal Study of Factors Which Are Predictive of Reading Success of Low Socio-Economic Children in the Seventh Year of School 202
Leslie McClellan

Conceptual Difficulty of Reading Material: Identifying Conceptual Constructs for Prediction 204
Robert J. Tierney and Joseph L. Vaughan, Jr.

Some Concurrent Validity Studies for Group Tests of Word Identification Skills 209
P. David Pearson, Brian Weeks, and Beulah Nuval

Modification of Pupil Attitude Toward Reading Through Positive Reinforcement Scheduling 219
Robert L. Aaron and Hal W. Seaton

Recognition of Sentences as a Function of Semantic and Syntactic Orienting Tasks 221
Fred W. Ohnmacht and James T. Fleming

A Study of The Influence of Minority Groups in the Selection and Development of Basal Reading Programs; A Five Year Replication 230
M. Jean Greenlaw

LANGUAGE

English Language and Concept Development Among Primary Level Hutterite Children – A Case Study in Intercultural Reading Education 232
J. Ray Stewart, Dorothy M. Lampard, and Myrna McGregor

Acquisition of Concrete and Abstract Words in a Beginning Reading Sample 240
George McNinch, Mark G. Richmond, and Sondra Kutzman

A Study on the Child's Conception of a Spoken "Word" 242
Jerry L. Johns

The Recognition of Word Boundaries by Adult Second-Language Subjects 243
David C. Johnson and Albert J. Kingston

Logic and Meaning of Language in Written Communication . . . 246
James A. Dinnan and Robert A. Lodge

An Oral Language Inventory for Remedial Readers: Grades 1 - 5 249
Joyce M. Hamrick

COMPREHENSION

Developing A Skills – Based Approach to Reading Comprehension 251
Wayne Otto

Identifying Sentence Meaning Skills 258
Karlyn Kamm

Identifying Passage Meaning Skills 259
Ginger Van Blaircom and Sandra White

Identifying Word Meaning Skills 260
Beverly Morrison

Variables in the Development of Mature Reading Strategies . . 261
Phil L. Nacke and James A. Culbertson

The Relationship of the Uniqueness of Prose Passages to the Effect of Question Placement and Question Relevance on the Acquisition and Retention of Information 268
Mark G. Richmond

A Model of the Comprehension of Poetry 280
W. John Harker

A Comparison of Syntactic Writing Maturity with Reading Comprehension 291
Norma Johnson

The Effect of Study Constructed Questions on the Comprehension of a Passage of Expository Prose 292
Ronald Schmelzer

Measuring Readability and Reading Ability in Order to Predict Comprehension 293
Ronald P. Carver

Analytic and Synthetic Processes in Concept Learning and Their Relationship to Reading Comprehension 298
Jocelyn Chen and Richard H. Bloomer

The Interactive Effects of Verbal Comprehension, Nonverbal Performance, Review Question Type, and Review Question Placement on Reading Comprehension 303
Randall A. Silverston, R. Jon Leffingwell, Don M. Beach, and James Benson

The Contribution of Tests of Semantic and Syntactic Context Clues to the Prediction of Two Reading Comprehension Measures 304
James W. Cunningham and Patricia M. Cunningham

Effect of Reader-Generated Questions on Learning from Prose 310
Jane M. Morse

Testing Comprehension of the Central Thought: Selecting Versus Generating Main Idea 317
Ginger Van Blaricom and Sandra White

Generation Versus Selection: Sentence Meaning 324
Karlyn Kamm

PRESIDENTIAL ADDRESS
WHO ARE NRC?

EDWARD FRY
Rutgers University

Maybe I should have said "Who is NRC?" but it is more fun to say, "Who are;" besides that, it has the kind of illiteration for emphasis on the "R" which is what we intend to emphasize.

Looking backwards is always easier than looking ahead so let us begin there. The National Reading Conference was founded as the Southwestern Reading Conference by a professor at Texas Christian University named Oscar Causey and a handful of young men who were his graduate students or who were friendly with them and were struggling with the current problems of reading. One of the problems that they addressed themselves to because of personal job assignments was "How do you improve the reading skills of college students?"

It is interesting that to this very day one prominent way that young professionals in the reading field get started is by teaching college reading improvement courses. In this current program, and in virtually every past NRC program, you will find papers being given on some aspect of this theme which varies from descriptions of successful programs to more research oriented evaluations of total programs and on to specific aspects of college level reading course applied methodology.

However, NRC has refused to be limited to college level reading partly because the field of reading refuses to be meaningfully divided by grade level and partly because most of those young men who surrounded Oscar Causey grew up to become professors of reading in many of the nation's prominent universities, and they were faced daily with the realities of teacher training and interests that led more directly into more basic reading research. In this year's program you will see papers continuing a long standing interest into the cloze procedure, measurement, linguistics, comprehension, systems, interests, and process.

On a more practical side we give and listen to papers on teacher training, reading failure, reading readiness, and adult literacy. It is dangerous to call some things "far out" but we have had papers on heart beats, poetry, and astrology, so we can't be accused of being too closed minded.

However, policy in the direction that an organization takes is not often thought out and planned; like many other institutions, it is shaped by decisions on specific instances. I remember one interesting and emotionally charged incident that occurred at this very meeting site several years ago when we had the revolt of the junior college instructors. As happens from time to time NRC was faced with a budget crunch and our then inventive president, Al Raygor, rose to the occasion by instituting a one-day pre-conference teacher training conference for college reading instructors. This pre-conference was widely advertised to junior and senior college deans and no small amount of them responded by sending neophyte college reading instructors. A certain percentage of these well meaning English literature instructors and upgraded public school teachers were perturbed that in one day they did not quite know enough to run a college reading program so they stayed on for the balance of the regular program. For some reason or other, what Wendell Weaver or Al Kingston were talking about did not seem particularly relevant and they help a rump meeting down the street to attempt to force the NRC program to get in better tune to their problems. We politely suggested that they join IRA which many of them subsequently did and now there is a special division of them which is about three times larger than NRC. I do not mean to imply that junior college reading instructors are not welcome in NRC, but we couldn't have their classroom-oriented tail wagging our research-oriented dog.

NRC has been similarly accused of taking a harsh line toward classroom reading teachers and public school reading specialists. Here again, policy has been made by practice. NRC again, partly for monetary reasons, used to give some sessions aimed at classroom teachers. In fact, the Saturday morning program has some real teacher pleasers on it when the third grade teachers could come in with the payment of a slight fee and say, "Gee, is that really George Spache?" Again, IRA has taken ample care of these kinds of conferences on the international, national, state, and local level. Any fair-to-middling reading specialist can go to a reading conference at least three out of every four weeks and see some "big names" without traveling much farther than the adjacent state.

Let us take a few moments then and see how NRC differs from other professional organizations. first of all, let me state that there is no hostility between IRA and NRC. Probably a majority of our members belong to both. Two IRA presidents have also been NRC presidents and many NRC members are active at all levels of IRA; however, our functions are quite different. As I see it, IRA is a giant professional organization that fills a variety of needs. It is first of all an in-service teacher training organization; a glance at its program and at its list of publications will amply demonstrate that. It is also a marketplace where authors and publishers try to influence school systems to buy their wares and again, I don't mean that this is bad as there is not much evidence that state or public agency–produced–curriculum materials in any way superior to private-enterprise produced materials. IRA also works at information dissemination and was particularly effective when they had the ERIC/CRIER at Indiana University but even without it, their continual outflow of monographs, journals, bibliographies and news letters helps to keep the field informed. If I would make any criticism, it is that their presses roll too often, but perhaps this is necessary if one views IRA as having a primary objective of teacher training.

Their efforts in the international area are laudable. Bringing together reading specialists around the world is a valuable link in intercultural friendship and information exchange.

IRA also has a research effort. Their national program solicits original papers and they have an excellent program of outstanding dissertation awards. Their *Reading Quarterly* is top quality. It is always dangerous to compare research quality, so I shall avoid it, but if one simply looks at the quantity of research papers given at the national conference on reading, NRC is far ahead of IRA, and all other reading organizations.

IRA has also made some progress in setting forth professional standards for reading specialists and in publishing a code of ethics. In fact, I am co-chairing a State IRA committee which is attempting to increase the number of required reading courses for elementary and secondary teachers. Prevously, the state committee aided in establishing legal minimum standards for reading specialists' certificates.

NCTE, The National Council of Teachers of English, is much like IRA, a real giant with many chapters and hyperactive printing presses, but it has a very broad range with listening, speech, writing, literature, linguistics and reading, kindergarten to graduate school.

IF NRC is passing itself off as a research-oriented organization, how then does it differ from American Psychological Association and others. First of all, the psychologists have a long and honorable history in reading research; we share such giants as Edward Thorndike, and many of today's reading authories, have basic psychological training. However, if you look at any recent programs for the APA or its regional meetings, you will find relatively few reading type papers. True, some of the verbal learning, psychometric, perception and other papers are directly related, but they are on a more indirect and abstract level. One could argue that psychological research is also directly related to reading success or failure, and this, too, shows us that we must draw the line somewhere just as they, too, draw the line as to how related their research might be to reading instruction.

I am reminded of a debate we had in New Jersey about the advisability of teaching reading in kindergarten in New Jersey. One speaker hoped to garner a little support by wiring Jerome Bruner and solicit his opinion about the advisability of early teaching of reading. Professor Bruner wired back, "Alas, I am not a reading specialist."

Our friends, the linguists, have made many contributions to the reading field;; they have changed our terminology and some of our concepts. "I didn't used to know what a grapheme were, and now I are one." However, when they attempt to apply their knowledge directly to reading, it sometimes seems like they haven't done their homework. I don't wish to depreciate the many scholarly efforts of the late Charles Fries, but when he developed a new linguistically-oriented reading series, it had a look strangely like McGuffey's and on application did not work any better than the much maligned traditional basal readers. If you were to look at the program of the Modern Language Association, to which most practicing linguists belong, many of their papers would be relevant to reading, but with about the same degree of relationship as verbal learning or physiology.

AERA probably comes closest to NRC in relevance and degree of sophistication of papers presented. However, AERA covers the total educational spectrum, general curriculum organization, school administration, and learning and measurement in all areas. Reading is there in a small special interest group, but it is almost overwhelmed by so many hundreds of other papers and thousands of conference attendees.

When looking at other organizations we shouldn't exclude the strong and viable college level reading associations like Western College Reading Association, College Reading Association, North Central College Reading Association, etc. . These organizations tend to have a regional character and they tend to be a little like IRA on the college level. While they have some excellent research papers, they have many members who are interested in college reading instruction problems.

It can be said that NRC also has a certain regional character. We seldom meet west of Texas, or north of Missouri, and we seem to like Florida, Georgia, and New Orleans rather well. Part of the problem is that it is easier to get people to go to Florida in December than to Minnesota and affording New York hotel prices is a bit of a chore.

Who then are NRC? We are a group interested in reading as a major lifetime activity. Most of us earn our living teaching reading and one way or another manage to do some research and/or writing about reading on the side, though occasionally we get lucky with grants or deans and get a little clear time for more research and development. Most of us teach in college or are graduate students in reading, but some of us are in the public schools. As an organization we are not competing with the giant IRA and we have a more specialized interest in reading than the psychologists, linguists, or general educational researchers.

As an organization, where are we going? I am happy to report that we are not going broke. Partially, courtesy of our friends at the ERIC/RCS Center who co-sponsored our yearbook, have helped us break even and our membership is stable. At Vital Issues and the Board of Directors we sometimes discuss growing, and I trust I am reflecting the current majority opinions when I say we do not want to grow in numbers of members. We do, however, passionately want to grow in quality of members. The field of reading needs the best researchers and practioners it can get. We certainly aren't going to kick anyone out, but we do need to replace the natural attrition with the best people in the nation. Professors have been helpful in encouraging some of their best graduate students to join and sometimes they remain. They are the leaders of tomorrow. Our porgram chairman, Jaap Tuinman, has encouraged some fine invited papers and your conversations with your colleagues can help us achieve this quality growth goal. We should all continually seek out top quality reading specialists in the public schools or in other fields like psychology or linguistics and specifically invite them to come to a few meetings.

NRC has a definite psychological approach to reading. Our members tend to be more at home with a test of significance than a Caldecott Award, but this doesn't mean that we shouldn't have a diversity of opinion around. Research study after research study has shown the importance of socioeconomic status on reading achievement, yet we have very few members with even minor sociological interest. We regularly give papers that would be accepted at linguistic meetings, yet we have few members with doctorates in linguistics. We tend to have very few papers dealing with motivation or the affective domain. We don't deny its importance but the research methods aren't too clean, so we tend to avoid them.

If we have a major conflict, it is probably how much we should emphasize the college and adult level instructional programs to the exclusion of elementary or secondary. You can look at this year's program to see how we are resolving this, or should I say what the balance is becoming. Personally, I feel equally at home in a first grade or a college freshman reading lesson. I have recently been advocating study skills, cloze, and reading rate improvement in an elementary teachers methods textbook. It seems to me a large number of papers here apply at all levels. If I might hazard a generalization, NRC tends to be a bit more applied at the college and adult level but researchers across the board. Whether or not NRC will remain in this posture doesn't depend on my opinion but rather the membership who largely build the program by submitting papers and symposia. I suppose the Board of Directors could order the program chairman to be more selective or to exclude certain levels, or the program committee could undertake such a definite policy, but so far it hasn't.

Whether or not NRC will outgrow its slightly regional character depends partly on the resulution of the above college level program. When we did go to Los Angeles some years ago we did not have a very good attendance. This may have been because the Western College Association was fulfilling the college level function.

And now if I might indulge in a few opinions about the more general state of reading:

The Right to Read Movement is finished. Hopefully, it won't die but be transformed into a more basic national and local effort to support reading research and development but the notion of having everybody literate by 1980 is simply not viable. Former U. S. Education Commissioner Allen had his heart in the right place, but he slept through his educational psychology class. You can't take the bottom 20% of the normal distribution curve and move it up to the 50th percentile or even chop off the curve at the 25th percentile and have a big bulge there, when you are dealing with a population the size of the United States. What you can do is to move the mean up a little bit and this in effect causes nearly everybody to read a little better, but it increases the standard deviation which means that both the bottom and top are farther away from each other, and both are farther away from the median. If through diligent research, and a lot of darn fool luck, we were able to come up with a super hot shot reading method for the bottom end of the curve, in all probability it would work just as well for the middle and top of the curve, and just how pray tell, would you be able to deny that method to all those screaming middle class mothers' children. Preschool enrichment only started to show a faint glimmer of promise, when they were demanding it for all children. The Electric Company was designed for inner city intermediate children and it is watched most dilligently by suburban primary children.

In any event, the funding amount is ridiculous. The current Right–To–Read annual budget of $17 million is less than one dollar for every elementary child. We might be able to make substantial headway with existing knowledge upgrading adult illiterates, but it would cost more like a battleship than a fighter plane.

If the government is going to make any significant effect in reading research, it has to stop changing its mind and funding sources every year or two. The Targeted Research was just getting started and "wham," it was chopped out of funds. The First Grade Studies might not have been the best research in the world, but they were pretty darn good compared to anything done before that. They, at least, had standard research design procedures, large populations, and consistent statistical procedures. As some of the methods were beginning to be refined and design errors corrected, the government gas ran out. We probably need a variety of long term research centers; some that do curriculum development like the successful Wisconsin Design; some that do large scale studies; and some that do and fund small studies.

The problem of models or systems continues to plague us, not because they aren't potentially valuable but because they are based on too few facts. Everybody, or at least a lot of us, want to be the Einstein of the reading field and have the model or system or theory. I don't think the field is ready for that yet. We need more facts. To get those facts we need better instrumentation. We are improving our paper and pencil reading achievement tests and that is all well and good, but we need a lot of other avenues; probably some of it psychbiological, such as EEG, EKG, EMG, etc. We have peered into eye movements about as hard as possible, but other than that, physiology is pretty much untapped. Motivation and personality research related to reading is all but a waste land. We have looked at language independent of people, but we haven't looked very much at environment–noise, air quality, food quality, home conditions.

Fortunately, those hard working teachers out there in the schools don't wait for the researchers to come up with the system; they are in there at eight o'clock every morning teaching reading by "conventional wisdom." It is possible that we can tinker with the system or give them a few more tools to get a little more efficiency out of it. Maybe improvement will be evolutionary, not revolutionary.

The National Reading Conference has a solid base upon which to build. Our yearbooks read little a scholarly study of the history of reading research for the past 25 years, not all of the history, of course, but such a significant amount that no future historian could ever omit us. We have a respected and established journal, *The Journal of Reading Behavior*, that is refereed and indexed. But our secret power is the glue that holds us together with common interests and mutual respect. We might be exclusive, but it is self-selective; anyone with a few dollars can join, and if you can stand listening to those boring papers, you are one of us. We are proud that we were racially integrated before some of the meeting places we selected, and though the hotel management frequently sees us as a bunch of characters who are always trying to sneak our own liquid refreshments onto their premises, we view ourselves as knights with shining textbooks out to slay the dragon of illiteracy.

A CRITICAL REVIEW OF RESEARCH AND EXPERT OPINION ON THE UNDERLINING STUDY AID

WILLIAM G. BROWNING
University of Minnesota

Over the years, questionnaire surveys have been used to investigate the study methods of students. For example, Charters (1925) had 258 women at a community college respond to a questionnaire on textbook study methods and other topics. Results indicated that only 4% of the students used underlining, compared with 38% who outlined and 35% who wrote marginal notes.

Recent surveys appear to support the view that underlining is a popular study aid. As part of her dissertation, McKune (1958) used a five point scale to find out about students' study methods. Underlining was found to be the most popular method, with 87% of the students saying that they either always or frequently underlined groups of words. Ritter (1971) described a study skills program that taught both the underlining and outlining study methods. Underlining was considered by a majority of the 232 students questioned to be the more useful skill.

Adams (1969) performed a study that provided some additional evidence about the popularity of underlining. Textbooks used by students in a reading education class and a social studies/science education class were collected three times during the term. Of the 60 students who owned their textbooks in the two courses, only two made no marks. The mean amount of underling over 10 pages of material was between about 65 and 70 inches. Adams concluded, "Underlining is practices by students in education courses as a study technique." In a survey of 200 randomly selected, used textbooks, Fowler and Barker (1974) found that about 92% of them had "significant applications" of emphasizing, including underling and highlighting.

Expert Opinion

It appears, then, that some kind of underlining is used by many of today's college students. Do study skills authorities generally share with students this positive attitude about underlining? According to Adams (1969), writers usually have not given underlinging "a place of significance in the hierarchy of skills." A survey by Laycock and Russell (1941) seems to support this assertion for earlier study skills books. The authors examined 38 how-to-study books written between 1926 and 1939. Only eight of the sources recommended underlining, compared with 21 that recommended outlining.

To update previously published reviews of how-to-study manuals, this writer reviewed 19 books and one journal article on the subject published between 1957 and 1975. The manuals reviewed were those on hand at the University of Minnesota Reading and Study Skills Center.

Table 1 shows the recommendations of the authors concerning five frequently mentioned study aids: recitation, underlining, marginal notetaking, outlining, and "combined" methods like SQ3R, OK4R, etc. An attempt was made to rank the recommendations of each author. A "1" means that the method was recommended with very few or no reservations. A "2" indicates that the method was recommended but was given secondary rank to methods marked with a "1" for that author. A "0" means that the method was not mentioned by the author, and an "X" shows that the author recommended against the use of the method. These rankings were made subjectively by this writer and are, of course, open to differences of opinion.

Table 1

Recommendations of Authors

Ratings

1 = Recommended with few or no reservations

2 = Recommended, but secondarily

0 = Not mentioned

X = Recommended against use

Key

R = Recitation, self-testing

U = Underlining

MN = Marginal notes

OND = Outlining, notetaking, diagramming

SQ3R, etc. = SQ3R and similar combined methods

Sources (in Chronological Order)	Date	R	U	MN	OND	SQ3R
Morgan & Deese (1st Ed.)	1957	1	2	0	1	1
Fedde	1961	1	X	1	1	0
Robinson	1961	1	2	0	1	1
Smith et al	1961	1	0	0	1	1
Pauk (1st Ed.)	1962	1	1	1	1	1
Resnick & Heller	1963	1	1	1	0	1
Hill & Eller	1964	1	0	0	0	0
Austin et al	1966	2	0	0	1	0
Gilbert	1966	0	0	0	1	0
Stanton	1966	1	0	1	0	0
Libaw & Martinson	1967	1	2	2	1	1
Anderson	1969	1	0	2	2	1
Kalish	1969	1	1	1	1	1
Morgan & Deese (2nd Ed.)	1969	1	1	0	1	1
Raygor & Wark	1970	1	1	0	1	1
Baker	1974	0	1	1	1	0
Pauk (2nd Ed.)	1974	1	1	1	1	1
Preston & Botel	1974	1	1	0	1	1
Policastro	1975	1	2	2	1	1

Table 2 is a summary of the recommendations found in the 19 sources. To a large extent, the generalization is borne out that underlining is less popular than some other study techniques. In fact, of the five study aids under consideration, underlining appears to be less popular than all the other aids except for marginal notetaking. For example, underlining received first preference less often than recitation, outlining, or SQ3R and other "combined" systems.

Table 2

Summary of Recommendations

	Recitation Recitation	Underlining Underlining	Marginal Notes	Outlining Etc.	SQ3R Etc.
First Preference	16	8	7	15	13
Second Preference	1	4	3	1	0
Not Mentioned	2	6	9	3	6
Recommended Against	0	1	0	0	0

It is interesting to note from Table 1 that there seems to be more acceptance of underlining in the newer sources. For example, among the more recent nine sources listed (1967-1974), six gave underlining a first preference, two gave it second preference, and only one failed to mention it. By contrast, among the earlier ten sources (1957-1966), only five mentioned the method, and of these, two gave it second preference and one recommended against using it. It is not possible, however, to say whether these figures are representative of study skills sources in general.

On key issues concerning underlining, there are varying amounts of agreement among the experts. Even where a general consensus exists, it seems to be based more on intuition than on any available research. For example, many of the authors discuss the advantages and disadvantages of underlining without having any research to support their conclusions. The authors generally agree that many students underline too much of their material (Fedde, 1961; Kalish, 1969; Policastro, 1975; Raygor & Wark, 1970; Robinson, 1961; Wark & Mogen, 1970). Other criticism are that underling tends to be "Hit or miss" (Morgan & Deese, 1957, 1969), confusing (Fedde, 1961; Robinson, 1961), and too easy or passive (Fedde, 1961; Robinson, 1961). Also, underlining is said to lose significance when the reader reviews the material later (Policastro, 1975).

On the other hand, some authors argue the opposite, saying that underlining helps to involve the student in the reading (Pauk, 1962 & 1974) and that it makes reviewing easier (Pauk, 1962 & 1974; Wark & Mogen, 1970; Preston & Botel, 1974). With one exception (Fedde, 1961), the authors allow that, when done right, underlining can be a useful study aid.

Although most of the authors agree that untrained students tend to underline too much, there is less agreement about what constitutes an "acceptable" amount. Policastro (1975) feels that even 20% is too much, for "it would result in 80 pages of rereading (while turning all 400) for review . . . ." On the other hand, Preston and Botel (1974) recommend that 15% to 20% of underlining is acceptable. Wark and Mogen (1970), in their workbook on underlining, take a less dogmatic position. Without giving a specific percentage, they state that the student should underline enough to highlight all the main ideas and supporting information. In the examples that they give, the average amount underlined is 30% and the range is from 18% to 42%.

Another point of controversy concerns what kind of underlining to do. For instance, some authors (Morgan & Deese, 1957, 1969; Wark & Mogen, 1970; Preston & Botel, 1974) recommend what might be called a "telegraphic" method, which involves underlining just enough words to get the essence of the passage. When reread, the underlined words make sense by themselves. On the other hand, Pauk (1962, 1974) suggests that the student only underline key words. In the examples he gives, the underlined words are not complete thoughts that can be read independently from the rest of the passage. Robinson (1961) suggests that stars, double underlinings, and similar devices be used to show degree of importance in a passage.

Many authors agree that a reader should not underline during the first reading of a paragraph or small section (Pauk, 1962, 1974; Robinson, 1961). Wark and Mogen (1970) go so far as to suggest that to keep from underlining too soon, the reader should not hold the writing instrument during the first reading. There also are some minor variations on the general theme of when to underline. Resnick and Heller (1963) recommend that the student read a section, then underline main ideas, and, lastly, go back to underline significant details. Morgan and Deese (1957, 1969) believe that during the first reading, the student should use checks or brackets; then, on the second reading he or she should underline after reading a couple paragraphs at a time.

Empirical Research

Many authors have lamented about the lack of empirical evidence behind most study skills recommendations (Berg, 1966; Adams, 1969;

Idstein & Jenkins, 1972). After a careful search of empirical studies on underlining, this writer has concluded that the underlining study aid is no exception to that complaint. In fact, there is little or no evidence to support any of the important issues discussed above.

The research on underlining can be divided into five main categories: (a) work which examines the isolation effect (von Restorff effect) when single words, syllables, numbers, etc. are isolated against a homogeneous background; (b) experiments on textual material where the experimenter does the underlining; (c) experiments where the textual material is underlined by the Ss in one or more of the treatment groups; (d) studies that include a comparison on experimenter- and subject-underlined treatments; and (e) experiments comparing subject-underlining with other commonly recommended study methods.

Wallace (1965) provides an excellent review of research on the isolation effect with non-textual material. In 1933, von Restorff reported a series of experiments indicating that an item is learned better when isolated against a homogeneous background. Subsequent studies have used three basic methods to investigate this phenomenon: isolation by making an item more vivid, such as by using a different color; isolation by including a different kind of item, such as a nonsense syllable among meaningful words; and isolation by manipulating the structural organization within a list. Studies of the first type are most closely related to our present concern with underlining as a study aid.

Based on the studies reviewed, Wallace (1965) concluded that isolated items appear to be learned more rapidly, but at the expense of other items in the list; so there is no overall difference in speed of learning. He also felt that the effect of isolation on retention could not be determined by the studies he reviewed because they did not control for differences in degree of learning.

Wallace made three other points of interest to our topic: First, there is some evidence that when items are isolated by being made more vivid, the learning of adjacent items ("spread effect") is somewhat improved. Secondly, the magnitude of the isolation effect appears to increase with increased vividness. And finally, the isolation effect seems more pronounced when an item is placed against a low meaningful background.

More relevant to our concern about underlining as a study aid is the research that involves textual rather than non-textual material. The first experiments to be considered are those in which the underlining of textual material was done by the experimenters. Christenson and Stordahl (1955) using Air Force basic trainees found no significant advantage in immediate or six-day delayed retention for underlining or for any of five other study aids. The criterion test reliabilities appeared to be adequate, ranging from .61 to .85. The design of the study, however, has some problems. Ausubel (1963) pointed out that the lack of significant differences may be partly due to leveling effects caused by (a) using familiar material, (b) using the same test as both a pretest and a measure of retention, and (c) testing the same Ss for both immediate and delayed retention.

Klare, Mabry, and Gustafson (1955) also used airmen (N = 939) in a study of the effects of underlining on immediate retention. Two kinds of underlining were used: (a) words were underlined which would appear in the correct answers of the criterion test, and (b) all important words were underlined. These two kinds of underlining were compared with a non-underlined version. The 50-item multiple-choice posttest had a split-half reliability of .87. The investigators found that neither underling method had a significant effect on the reading retention of the Ss. More importantly, an ANOVA revealed a significant interaction between Ss' mechanical aptitudes and test scores. Ss with high mechanical aptitudes were helped by the underlining while those with low aptitudes were hindered.

Hershberger (1964), in a pre-posttest study of fifth graders, found no significant improvement in immediate or delayed retention from use of a complex cueing system that included underlining, variation of type size, and variation of color. The use of gain scores (pretest scores subtracted from posttest scores) is a weakness in this study.

Cashen and Leicht (1970) investigated the effect of underlining on the retention of materials read by college freshman (N = 40). No information was given on the reliability of the 15-item multiple-choice test. Ss in the underlining condition scored significantly higher on items covering both underlined principles and adjacent material than did Ss in a non-underlining control group. The effect was consistent on three different passages.

In a later study, the same authors (Leicht & Cashen, 1972) used four different treatments: no underlining, underlining of principles, of examples, or of trivia. While there were no significant differences in overall scores, Ss in each group scored slightly higher on the subset of test items dealing with the underlined material for that group. The authors' earlier finding (Cashen & Leicht, 1970) that showed an effect on adjacent material was not replicated. The authors did not report on the reliability of the criterion test.

Rocchio (1972) divided 21 high school students into two groups, one receiving an underlined 2000 word passage and the other a non-underlined version. Multiple regression analysis found no significant differences in test scores between the two groups. There was, however, a trend favoring the underlining group. No interaction was found between reading ability and treatment. The somewhat low reliability of the criterion test (.53) and the very small sample size are weaknesses in the study.

Crouse and Idstein (1972) performed two studies on the effects of experimenter underlining. In the first study, four groups of college students were compared: (a) a control group which read a non-underlined passage; (b) a group which studied an underlined passage after being told that the test would cover the underlined parts; (c) a group with the same material as group 2 but told to think of what questions might be asked of the underlined portions; and (d) a group for which the actual questions on the test followed the relevant underlined portions. Although mean scores for the underlining groups were slightly higher than the control, the differences were non-significant. One possible cause was the potential leveling effect from using very short passages totalling about 600 words.

The second experiment used a much longer (6000 word) passage and involved two conditions, a non-underlined control and a group which knew that the underlined portions of their material would be covered on the test. The mean score advantage of the underlining group was exceptionally high: 17.30, versus .924 for the control group (significant beyond the .01 level). The unusually large differences in mean scores may be partly explained by the fact that the Ss in the underlining group knew that all the test items would come from the underlined portions of the passage. Results indicated that Ss classified as fast- and medium-learners derived greater benefit from the underlining than did slow-learners.

Fowler and Barker (1974) conducted two experiments on underlining and "highlighting." The first experiment will be discussed in a later part of this review. In the second experiment, portions of the material were matched according to difficulty level so that, in a counterbalanced design, each S had some parts highlighted, some parts underlined, and some parts without any emphasis (control). Performance on items from underlined material was significantly better than for the control condition. The performance for the highlighted condition was not significantly different from that for the other two conditions. The results indicated that for experimenter-emphasized material, a highlighter is no more effective than a standard red pen.

In an early study by Dynes (1932), reading and rereading textual material was compared with a study method that included a preview, a detailed reading while underlining, a review, a written summary, and a recitation. Using gain scores in a pre-posttest design with 134 high school students, the author found that there was a slight, statistically significant advantage in three-week delayed retention favoring the six-step study method.

Using college freshmen, McKune (1958) had the Ss in the treatment group (N = 175) underline the words or phrases which they considered important in a 2 page passage. Ss in the control group (N = 184) read the passage without underling it. Immediate recall was measured by a 26 item multiple-choice test with a split-half reliability of .85. The same test was given to the Ss by surprise 3 weeks later to measure delayed retention. No significant differences were found for either testing. The lack of significant differences might be partly explained by shortness of the passage.

Idstein and Jenkins (1972) conducted 2 experiments on the effect of underlining by college Ss. The first was a 2 X 2 ANOVA posttest-only design comparing underlining with repetitive reading and a 9 minute review time with a 4½-minute time. The Ss reviewed the 1200 word passage 1 week after the initial contact and then immediately took a 24 item completion test. There were no significant test score differences between the reading conditions, nor was there any significant interaction between review condition and reading method. There was, however, a significant difference favoring the 9-minute review group over the 4½-minute review group.

In a second study reported in the same article, the authors chose a longer, 6000 word passage which they believed approximated more closely the material college students study. Ss in the underlining and repetitive reading groups read the passage during one session and reviewed it and took the completion test during a second session two weeks later. No significant difference was found in their scores.

Adams (1969) investigated the correlation between amount of underlining done by college students and their course grades. Students' textbooks in a reading education course and in a social science/science education course were collected and tabulated three times during the quarter. The students were not told the reason for the collecting. Correlations of .20 and .51 were found for the reading and the social science/science courses, respectively. The study was based on the assumption that students who underline more would tend to receive higher grades. But if, as some experts believe, many students underline too much material, then the low-to-moderate correlations might be partly explained by this fact.

Four studies were found that included a comparison of experimenter- and subject-underlined material. Earp (1959) randomly assigned 144 college students to three groups. One group read and underlined two selections from a psychology textbook, another group read a pre-underlined version, and the third group read a non-underlined version. A multiple-choice test used to measure immediate and delayed retention failed to show any significant differences.

Schnell and Rocchio (1974) reported two studies, one of a high school population and the other of a junior college population. In each study four groups were used: (1) receiving 110 minutes of instruction and practice in the underlining method suggested by Wark and Mogen (1970), after which the Ss read and underlined the 1100-word passage; (2) without previous instruction, reading a version of the passage that was underlined by the experimenters, again using the Wark-Mogen method; (3) reading and underlining without prior instruction or practice; and (4) reading a non-underlined version (control). A 20 item test with a split-half reliability of .70 was used to measure immediate and 11 day delayed recall. No reviewing was allowed before the delayed test.

All three treatment groups scored significantly higher than the control group on immediate recall. On delayed recall, a non-significant trend favored the underlining groups. Also, both of the subject-underlined groups scored significantly higher than the experimenter-underlined group on immediate recall but not on delayed recall. The group of Ss that underlined without prior instruction scored significantly higher than the group that received instruction. There was a significant interaction between reading ability and criterion test scores favoring the more proficient readers for all groups on delayed but not on immediate recall.

In the first of two experiments reported by Fowler and Barker (1974), Ss were randomly divided into four groups: (a) read and highlight material (ACL); (b) read material highlighted by a "yoke-mate" in the first group (PHL); (c) read material highlighted by the experimenters (EHL); and (d) read material without any highlighting (control) (NHL). All Ss spent one hour studying the articles, which totalled 8000 words, and then took the 54 item criterion test one week later after a ten minute review.

Although no significant differences were found among total scores for the four groups, the EHL group did significantly better than the NHL group on the items highlighted in the former group. The AHL group performed significantly better on items that covered material they had highlighted than they did on items from non-highlighted material. Also, the AHL group performed significantly better than their passive yoke-mates (PHL) on items that covered highlighted material.

Willmore (1966) found some significant differences in a comparison of reading, underlining, outlining, and SQ3R. Eighty college students in how-to-study classes received instruction in the techniques and then applied them to different chapters of an English history text. The multiple-choice tests used as criterion measures has split-half reliabilities ranging from .77 to .87. Analysis of covariance, with study time as the covariate, indicated that Ss scored significantly higher on the average when underlining than when using the other methods. Also, underlining took less time than the other methods.

Holmes (1972) used 41 students in a university study skills program to compare the effects of SQ3R and underlining. The students were divided into two treatment groups. Members of each group listened to a 15 minute tape recording about one of the study methods and then had six hours of practice. The underlining method taught was the "telegraphic" method recommended by Wark and Mogen (1970). The

Ss then spent two hours applying the method to a 5000 word passage on the history of Japan. Two days later they were given ten minutes to review the material, after which they took a 50-item objective exam. With reading ability controlled as a covariate, multiple regression analysis found a non-significant trend (.1388) favoring the underlining group. The lack of a control group in this study seriously limits any interpretation of the results.

Discussion of the Research

A **post-hoc** discussion such as this should be approached with caution. It often is easy to explain away non-significant results by pointing to weaknesses in the studies, while ignoring similar weaknesses in studies that show some significant differences. Therefore, some of the following remarks are intended only as hypotheses about possible contributing causes of some results.

Twelve studies were found that included a comparison between a control group and a group reading material in which the experimenter(s) had employed some kind of emphasis technique. Of these studies, six obtained significant results on some kind of retention test. No significant differences and/or interactions were found on the main variables in six other experiments. Some kind of trend favoring underlining was found in two of these studies.

Based on the studies reported, a few tentative conclusions can be drawn. First, in all cases except the second Crouse and Idstein (1972) experiment, any significant mean differences were relatively slight, on the order of about one or two items more for the experimental group than for the control. However, since none of the experiments came close to approximating the amount of material a student would have to remember for a college test, it is not clear whether this result can be generalized to college studying.

One partial explanation for the general lack of large differences may be the presence of an interaction between subjects' ability and the effect of underlining. Klare, Mabry, and Gustafson (1955) found that underlining helped high-mechanical ability airmen and hindered low-mechanical ability airmen. Crouse and Idstein's second experiment (1972) determined that fast- and medium-learners benefitted from the underlining more than did slow-learners. If such an interaction was operating in the other experiments, it could have cancelled out some of the main effect differences.

Eleven studies were found that compared the retention effects of a no-underlining control with the effects of underlining done by the subjects. Four of these experiments found significant mean score differences favoring the underlining groups. The Fowler and Barker (1974) study, however, found significant differences only on sub-scores; not on differences between total scores.

Seven studies reported non-significant differences between groups under different underlining conditions. Two of these experiments combined marginal notetaking with underlining in the experimental groups; it appears that under some conditions underlining while reading seems to improve retention and under other conditions it does not. Passage length and presence or absence of instruction in underlining may cause these differences.

For the studies under consideration, however, there is no obvious relationship between passage length and criterion test results. Among the four studies with significant differences, two used relatively short, 1100 word passages and two used longer passages (8000 words and chapter-length). Among the studies that did not have significant results, the lengths ranged from 1200 words to 6000 words.

Willmore (1966) provided instruction in each of the study methods investigated. Schnell and Rocchio (1974) used a group with and a group without instruction in each of their two studies. Both groups did significantly better than the control, and the non-instruction group did better than the instruction group. Fowler and Barker (1974) did not provide any instruction, but the test score differences in that study were very low.

While the possible effects of underlining instruction are not clear-cut, it seems at least plausible that lack of instruction in underlining might have helped to produce the large proportion of studies with non-significant results. In their discussion, Crouse and Idstein (1972) hypothesized that when untrained Ss underline textual material, they may simply underline the parts they would have studied anyhow, while ignoring other parts from which test questions are made. If true, this hypothesis might help to explain why significant results are obtained more often when the experimenters do the underlining than when the students do.

With very few exceptions, authors of underlining studies say little or nothing about the kind of underlining that they used or taught. The assumption in many cases seems to be that it is sufficient to simply say that underlining instruction was given. Yet, as we found from reviewing how-to-study manuals, there is some disagreement about how material should be underlined. Therefore, when no information is given about the kind of underlining being used by the experimenter and/or taught to the Ss, it is difficult to interprete the results of a study.

Suggestions for Research

As with other study aids, there are numerous questions about underlining that could be investigated. For example, the survey of how-to-study publications found that the following questions often were discussed:

1. Does underlining help retention?
2. Does underlining facillitate review?
3. What range of underlining amounts is optimum?
4. What kind of underlining is most effective? (E.g., key word, telegraphic, etc.)
5. When should the underlining be done?

In addition, some other important questions generally were not considered by the authors. For instance:

6. What effect does individual differences have on results from using underlining?
7. Does the effect of underlining vary with the kind or difficulty of the material being studied?
8. What rationale should the reader use for deciding what to underline? E.G., should all main ideas and details be underlined or just those considered likely to be on an exam?

Research on underlining mainly has dealt with the first question listed above and has lightly touched on a couple of others. The tentative answer to the first question seems to be that if the underlining accurately reflects the points covered on the test, then it is likely to improve mature students' test scores.

Generally, then, with the exception of question one, the other questions listed have not been dealt with in a systematic way, at least not in the studies covered by the present review. Until an effort is made to investigate these issues, study skills instructors and authors will continue to rely on anecdote, opinion, and intuition in recommending how, when, and whether students should underline their textbooks.

References

Adams, E.K. Underlining: The graphical aid to college reading In G.B. Schick and M.M. May (Eds.), *Reading: Process and Pedagogy.* 19th NRC Yearbook, Vol. 1. Milwaukee, Wisconsin: National Reading Conference, 1969.

Anderson, B.D. *Introduction to college.* New York: Holt, Rinehart and Winston, 1969.

Austin, L., et al. *College reading skills.* New York: Alfred A Knopf, 1966.

Ausubel, D.P. *The psychology of meaningful verbal learning.* New York: Grune and Stratton, 1963.

Baker, W.D. *Reading skills.* (2nd ed.) Englewood Cliffs, New Jersey: Prentice-Hall, 1974.

Berg, P., & Rentel, V. Improving study skills. *Journal of Reading,* 1966, *9,* 343-348.

Cashen, V.M., & Leicht, K.L. Role of the isolation effect in a formal educational setting: Underlined and nonunderlined printed material. *Journal of Educational Psychology*, 1970, *61*, 484-486.

Charters, J.A. How two-hundred and fifty-eight junior college women study. *Journal of Educational Research*, 1925, *11*, 41-49.

Christenson, C.M., & Stordahl, K.E. The effect of organizational aids on comprehension and retention. *Journal of Educational Psychology*, 1955, *46*, 65-74.

Crouse, J.H., & Idstein, P. Effects of encoding cues on prose learning. *Journal of Educational Psychology*, 1972, *63*, 309-313.

Dynes, J.J. Comparison of two methods of studying history. *Journal of Experimental Education*, 1932, *1*, 42-45.

Earp, N.W. The effect of certain study conditions on the achievement of college freshmen. Unpublished doctoral dissertation, Colorado State College, 1959.

Fedde, N.A. *Preparing for college study*. New Haven, Conn.: Readers Press, Inc., 1961.

Fowler, R.L., & Barker, A.S. Effectiveness of highlighting for retention of text material. *Journal of Applied Psychology*, 1974, *59*, 358-364.

Gilbert, D.W. *Study in Depth*. Englewood Cliffs, New Jersey: Prentice-Hall, 1966.

Hershberger, W. Self-evaluational reading and typographical cueing: Techniques for programming self-instructional reading materials. *Journal of Educational Psychology*, 1964, *55*, 288-296.

Hill, W., & Eller, W. *Power in Reading Skills*. Belmont, Calif.: Wadsworth Publishing Company, 1964.

Holmes, J.T. A comparison of two reading-study methods for college students. Unpublished doctoral dissertation, University of Northern Colorado, 1972.

Idstein, P., & Jenkins, J.R. Underlining versus repetitive reading, *Journal of Educational Research*, 1972, *65*, 321-323.

Kalish, R.A. *Making the most of college: A guide to effective study*. (2nd ed.) Belmont, Calif.: Wadsworth Publishing Company, 1969.

Klare, G.R., Mabry, J.E., & Gustafson, L.M. The relationship of patterning (underlining) to immediate retention and to acceptability of technical material. *Journal of Applied Psychology*, 1955, *39*, 40-42.

Laycock, S.R., & Russell, D.H. An analysis of thirty-eight how-to-study manuals. *School Review*, 1941, *49*, 370-379.

Leicht, K., & Cashen, V. Type of highlighted material and examination performance. *Journal of Educational Research*, 1972, *65*, 315-316.

Libaw, F.B., & Martinson, W.D. *Success in college*. (2nd ed.) Glenview, Ill.: Scott, Foresman and Company, 1967.

McKune, E.J. An investigation of the study procedures of college freshmen. Unpublished doctoral dissertation, Colorado State College, 1958.

Morgan, C., & Deese, J. *How to study* (1st ed.) New York: McGraw-Hill Book Company, 1957.

Morgan, C., & Deese, J. *How to study* (2nd ed.) New York: McGraw-Hill Book Company, 1969.

Pauk, W. *How to study in college.* (1st ed.) Boston: Houghton Mifflin Company, 1962.

Pauk, W. *How to study in college.* (2nd ed.) Boston: Houghton Mifflin Company, 1974.

Policastro, M. Notetaking: The key to college success. *Journal of Reading,* 1975, *18,* 372-375.

Preston, R.C., & Botel, M. *How to study.* Chicago: Science Research Associates, Inc., 1974.

Raygor, A.L., & Wark, D.M. *Systems for study.* New York: McGraw-Hill Book Company, 1970.

Resnick, W.C., & Heller, D.H. (Eds.) *On your own in college.* Columbus, Ohio: Charles E. Merrill Books, Inc., 1963.

Ritter, J.H. University study skills program. *Journal of Reading,* 1971, *14,* 377-380.

Robinson, F.P. *Effective study.* (revised ed.) New York: Harper and Row, 1961.

Rocchio, D.J. The effect of prior underlining on immediate comprehension of reading passages. Unpublished independent study project, University of Missouri at St. Louis, 1972.

Schnell, T.R., & Rocchio, D.J. A comparison of underlining strategies for improving reading comprehension and retention. Paper presented at the meeting of the National Reading Conference, Kansas City, Missouri, 1974.

Smith, D.E.P., et al. *Learning to learn.* New York: Harcourt, Brace and World, Inc., 1961.

Stanton, T. *R.S.V.P.: A dynamic approach to study.* Glenview, Ill.: Scott, Foresman and Company, 1966.

Wallace, W.P. Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin,* 1965, *63,* 410-424.

Wark, D.M., & Mogen, M. *Read, underline, review.* New York: McGraw-Hill Book Company, 1970.

Willmore, D.J. A comparison of four methods of studying a college textbook. Unpublished doctoral dissertation, University of Minnesota, 1966.

LEARNING FROM TEXT: A REVIEW OF THEORIES, STRATEGIES, AND RESEARCH AT THE HIGH SCHOOL LEVEL

HARRY SINGER
University of California, Riverside

ALAN RHODES
Upland High School

Overview

This review begins with a brief history of landmark research in the psychology of reading at the high school level. Together with references to other reviews of high school reading, it puts our current review into historical perspective. We than state some recent developments in high school reading research, and culminate our introduction with a separation of reading development into two phases: (a) reading acquisition behavior and (b) learning from text.

The review itself is organized into six parts. The first part explains three major problems in high school reading and learning from text which still need to be solved: (a) the range of individual differences in reading, and strategies for meeting this wide range in reading achievement, (b) attitudes toward teaching reading, and suggested antidotes, and (c) measurement of reading and learning from texts in content area. The second part consists of research on strategies for teaching students to learn from texts. It is subdivided into six categories: (a) advance or cognitive organizers, and structured overviews, (b) adjunct questions (questions inserted into text), (c) visual imagery in comprehension, (d) organization of passages and sentences, (e) contextual clues, and (f) reading guides. The third part draws upon implications from research and instructional theories for ways of enhancing students' abilities to learn how to learn from text. This section includes instruction in study skills. The fourth part reviews readability research, with stress upon handling individual differences by matching texts to students' levels of ability. The fifth part evaluates effectiveness of reading acquisition programs. The sixth part, on theories of reading, points out the limitations of general theories of reading, and summarizes a more appropriate theory: an information processing model. The review concludes with a brief appraisal of high school teaching and research, as well as a prediction that school-wide programs will increasingly emerge, enabling students throughout the range of individual differences to develop toward independence in reading and learning from texts in content areas at the high school level.

Landmarks in a century of research in reading. Inquiry into reading at the high school level has been pursued from the beginning of research into the psychology and pedagogy of reading. The first to measure eye-movement behavior, Javal (1879) found that mature readers moved their eyes in saccades or jumps from one fixation point to the next. Mature readers also perceive words more rapidly than their individual letters (Cattell, 1886). Huey (1908) discovered that as readers mature they shift from shorter to longer perceptual units and develop automaticity in processing these longer units. After analyzing errors made in reading a paragraph, Thorndike (1917) defined reading as reasoning. In a cross-sectional study of average readers at each grade level from 1 to 12, Buswell (1922) computed that high school seniors, in contrast to first graders reading the same passage, have relatively large spans of perception, shorter pauses, and more rhythmical eye-movements. High school seniors also varied their eye-movement behavior according to their purposes, as well as the difficulty of the material and the kind of material (Judd and Buswell, 1922). Moreover, speed and comprehension of reading were found to be different functions and, therefore, had to be assessed separately (Gates, 1921). But, the level of reading comprehension in high school is not uniform from one subject area to another (Bond, 1938), which means that students are likely to have uneven profiles of reading comprehension when tested in the various content areas. Indeed, success or failure in reading is a linear, weighted combination of multiple factors (Monroe, 1932).

High school readers are not only quantitatively but also qualitatively different from beginning readers. For example, high school readers have already attained functional oculomotor efficiency, while beginning readers still have to integrate internal and external mechanisms for controlling their eyes in reading (Gilbert, 1953). Also, the quarter of a second fixation pause in mature readers divides into stabilizing, seeing, and processing time (Gilbert, 1959). However, in some aspects of reading, such as judging relevancy of information and maintaining a set purpose in searching for information from text, little developmental change occurs over grades four to college (Gans, 1940), perhaps because systematic instruction in reference reading has not been given over these grades.

Nevertheless, individual patterns of reading vary at the high school level (Strang, 1942). For example, skills for prose are different from non-prose reading, and these types of reading require different subskills, abilities, and attitudes (Hall and Robinson, 1945). But, regardless of content or task requirements, readers do not just mobilize their strengths and minimize their weaknesses; all readers have to draw upon at least a minimum of certain decoding abilities if they are to read at all (Holmes and Singer, 1961, 1966). In general, as readers develop in speed and power of reading, their individual differences decrease in syntactic and phonemic components, but increase in reasoning, semantic, and morphemic subsystems for attaining speed and power of reading (Singer, 1965). Personality factors also account for individual differences in general reading ability (Atheny and Holmes, 1969).

Thus, research in reading at the high school or mature level of reading is not a recent focus of investigation.

Recent developments in high school reading research. What is new is that educators, researchers, theorists, and even funding agencies have begun to concentrate their training programs, modes of inquiry, ideational activities, and financial resources, respectively, upon reading comprehension beyond the reading acquisition stage of development. Some nine states now require all high school teachers to have a preservice course in teaching reading (Estes and Piercey, 1973; Freed, 1974). Although sparse, some research has been done at the high school level (Karlin, 1969) and in the content areas (Singer, 1970). Moreover, applicable to the high school level of reading comprehension (Goodman, 1976, Holmes and Singer, 1966; Ruddell, 1976), while specific theories and models on processing and acquiring information from text and beginning to be formulated (Bormuth, Carr, Manning and Pearson, 1970; Frase, 1968b; Miller, 1975; Rothkopf, 1966). Furthermore, the National Institute for Education began in 1974 to request proposals for research in post-primary comprehension. Implicit in this request is the recognition that reading development must be separated into two phases.

Separation of reading development into two phases. Two overlapping curves of reading development exist. These curves can be identified as reading acquisition behavior and learning from text (Singer, 1976).[1] Only recently has it become clear that reading acquisition behavior or learning to read consists of closed objectives, such as grapheme-phoneme correspondences, which can be mastered by 95 percent of the population (Bloom, 1971). Therefore, individual differences in this developmental phase of reading diminish as students progress through school. Consequently, reading acquisition behavior has a **decreasing** relationship with mental age as a group of students progress through school. In contrast, learning or **gaining** information from text consists of open-ended objectives, such as vocabulary ability, which are never completely mastered (Cronbach, 1971). Indeed, individual differences in open-ended objectives increase with concomitant progression through school. Hence, individual differences in acquiring information from text has an **increasing** relationship with mental age as a group of students progress through the grades (Singer, 1976).

Although some high school students, probably no more than five percent, are still learning to read, most students at this level have already mastered the reading acquisition stage of development.

However, even students who are beyond the general reading acquisition stage still need to learn how to pronounce new words, particularly technical terms, and, of course, they must also learn the meanings and syntactic properties of these words. What these students have mastered is ability to (a) integrate semantic, syntactic, and phonological aspects of their oral language ability with acquired rules for responding accurately to almost any graphemic stimulus, (b) associate meaning to printed stimuli, (c) test the resulting meaning to determine if it meets their expectancies for cumulative meaning, (d) integrate and store meaning with previously processed ones, and (e) encode the decoded message (Goodman, 1976, Singer, 1970). Since most high school readers have already mastered reading acquisition behavior, we will concentrate this review upon theories, strategies, and research on learning from text, emphasizing mostly those high school studies that have been reported over the past five years.[2] We shall start this review with the identification of three major problems and ways of solving them.

Major Problems in High School Reading and Learning from Texts

Range of Individual Differences in Learning from Text

Components of reading acquisition behavior and learning from text are difficult to separate, but an approximation for doing so, through repeated testing over grades one to four on a primary grade test until mastery has been attained by students, has been proposed (Singer, 1973, 1976). Although the distinction is important for differential diagnosis, it is not necessary for most high school students because they have already mastered the initial phase of reading development. Only students scoring in about the lowest 25th percentile of standardized reading survey tests need be given additional tests to determine whether they have not yet mastered the reading acquisition stage (Singer and Rhodes, 1977).

Results of reading survey tests indicate that the range of variability or individual differences in reading comprehension increases with progression through school (Cook, 1951). In general, variability in a heterogeneous group of students, ranging from an IQ of 67 to an IQ of 132, is two-thirds of the average age of the group (Bond and Tinker, 1957). Consequently, tenth grade reading ages will vary from an age-equivalent of 10 to 20 years. At grade 12, the range will increase to 12 years. This increase in range of variation will be true even if all individuals in the group are reading up to their mental age levels. Yet, some teachers, unaware of this expected range of individual differences and unprepared by their professional education for it (Otto and Smith, 1969), assume their students know how to read and understand their assigned textbook material (Herber, 1970b), or that students who cannot do so have reading difficulties that should have been overcome in junior high school or in elementary school. Apparently these teachers confuse a difference in ability with defectiveness, deficiency, or disruption in reading and learning from texts (Wiener and Cromer, 1967). They need to become informed about the normal range of individual differences. If they are then to provide equality of educational opportunity for the range of individual differences (Coleman, 1968), they must also have appropriate teaching strategies and materials (Singer and Donlan, 1977; Singer and Rhodes, 1977).

Strategies for meeting the wide range of individual differences in reading achievement. Traditionally high schools have attempted to handle the wide range of individual differences in reading achievement by narrowing the range of a class's reading achievement through tracking or ability grouping. However, this strategy still leaves groups heterogeneous on variables not used in the tracking criterion. It also results in inequality in educational opportunity through downward adaptation of curricula and tends to stigmatize students (Balow, 1962, 1964). These negative consequences can be avoided by using strategies for handling individual differences without formal groupings or tracking (Singer, 1973). Among the strategies are: learning to read guides; reading, reasoning, and evaluation guides (Herber, 1970b; Singer, 1973); active comprehension (Singer, 1973); the project method of teaching reading in content areas (Singer, 1973; Smith, 1965a); multilevel texts (Singer and Donlan, 1976); cross-ability teaching on a self-selective basis (Singer, 1973); directed reading activities (Burmeister, 1974); and use of a text as a data base for inquiry procedures (Ryan and Wheeler, 1974). These approaches presume teachers have skills for using them or attitudes for acquiring and putting them into practice.

Attitudes and Antidotes

Attitudes of high school teachers on teaching reading vary considerably. Some teachers presume all aspects of reading should have been taught in elementary school. Secondary teachers hold this attitude more strongly than junior high school teachers (Otto and Smith, 1969). A possible hypothesis is that secondary teachers react negatively to the undifferentiated term, "reading," with its implied instructional responsibility which asserted in an old slogan that "every teacher is a teacher of reading." A more favorable response from high school teachers might be elicited by a new slogan that teaching students to **learn from texts** in their content areas is the responsibility of all content area teachers.[3]

A need for more investigation and experimentation in attitudes is apparent. For example, Otto and Smith (1969), after constructing a reliable scale for assessing attitudes of content teachers, administered it to a group of secondary teachers. To the surprise of Otto and Smith, most of the secondary teachers revealed they had **favorable** attitudes toward teaching reading in content areas, but believed reading specialists were better prepared for doing so.

A compromise solution to the problem posed by this attitude is to have content reading specialists, those trained in both content and reading instruction, teach their colleagues how to teach reading in content areas (Cassidy, 1973; Singer, 1973). If only one reading content specialist is available to the school, then content teachers can communicate knowledge and mode of inquiry in their content areas to the content reading specialist who, in turn, can use this knowledge to demonstrate skills and techniques for teaching students how to learn this content and mode of inquiry from text (Hafner, 1974; Herber, 1970a; Robinson and Smith, 1973). The purpose of the school reading content specialist then is to move teachers along the continuum from frustration and complaints to experimentation and, finally, problem solving on teaching students to read and learn from texts. In other words, the goal is to move teachers from assumptive to prescriptive teaching (Ronbinson and Smith, 1973).

Another problem is measurement of ability to learn from text in general reading and in content areas. If learning from text is to be content-specific, it would be necessary to first differentiate general from specific reading tasks in content areas, then select or construct reliable and valid tests for assessing specific reading tasks, and, finally, use differences between an expectancy criterion and specific reading task performance that are statistically significant for diagnosing ability to learn from text in a specific content area.

Measurement of Ability to Read and Learn from Texts

Differentiation of general from specific reading tasks in content areas. Reading tasks can be defined and analyzed by different statistical models, each with its own basic assumptions (Holmes, 1965). Spearman's basic assumption is that all mental tasks and abilities draw upon two factors in varying amoungs: "g" is a general factor which correlated to some degree with all mental tasks and "s" is a specific ability which does not correlate with all other tasks but may overlap with other specific factors and with "g." Figure 1 depicts these relationships.

The Spearman model can be used conceptually to differentiate reading tasks in the content areas from general reading tasks. This differentiation can be accomplished by recognizing that reading tasks in content areas have varying amounts of "g" and "s." For example, reading for the main idea is a task common to all types of reading; hence, reading for main ideas is a general reading task. In contrast, interpretation of charts, graphs, maps, symbols, devices, and technical vocabulary can be categorized as specific reading tasks. Thus, reading in each content area may have both "g" and "s" factors. Consequently, teachers need to be cognizant of both "g" and "s" reading tasks in their content areas in order to determine which students need "g," "s," or both types of instruction. Teachers should also be aware of "g" and "s" overlap in standardized tests of reading achievement.

Overlap between general and specific tests.[4] General reading ability is assessed by such tests of comprehension, speed of reading, and general vocabularly, such as the Gates-McGinitie Reading Tests (Gates and McGinitie, 1965). Content for general or survey reading tests is drawn from different areas of the curriculum, and test questions tend to tap all levels in the hierarchy of reading objectives (Barrett, 1968), including details or literal statements, inferences, interpretations, applications, main ideas, and central thoughts.

Some content-specific tests focus on knowledge and processes of learning from text in a particular curricular area, such as literature and literary devices (Davis, 1960). Other content-specific tests consist of content in social studies, science, mathematics, and literature (Gardner et al., 1966; Tiegs and Clark, 1963), art, geology, history, and fiction (Robinson and Hall, 1949), natural sciences and social sciences (American Council on Education, 1970), library sources and skills (Shores and Moore, 1940), and use of the dictionary (Spache, 1963).

Both general and specific abilities are assissed in some test batteries. For example, the Iowa Silent Reading Test (Green, Jorgensen, and Kelly, 1942) includes not only general reading ability tests, such as rate and comprehension, but also such specific tests as poetry comprehension, use of an index, and selection of key words.

Reliability. Although the total scores on general and specific tests tend to have high reliabilities, usually about .80, their subtests frequently do not have adequate enough reliabilities to place much confidence in them, let alone use them for diffential diagnosis. If subtest reliabilities are low, reliability of differences between their scores, which is a prerequisite for a valid diagnosis, would, of course, be even lower (Mosier, 1951).

A major reason for low subtest reliability is that test producers try to limit tests to 45 minutes so they can be administered in a single sitting, one school period. This time limit requires subtests to be too short to reliably sample a domain, whether the domain is in levels of comprehension or in specific content areas. Short tests are limited also in capacity to overcome temporary interferences, such as attentional distraction, particularly when timed tests are used.

Validity. Although a general achievement test that samples content from all curricular areas may be reliable enough for generalizing test results to content areas, test users should content analyze tests to determine whether they are valid for a school's curriculum. If tests do not have curricular validity, they cannot properly be used for assessing what students learned in the curriculum. Hence, new tests have to be constructed, as Jewell (1969) did for a school district in Alberta, Canada, based on tests used there.

But, Marks and Noll (1967) defined validity of comprehension tests differently. They argued that reading tests should assess students' ability to gain information only from reading test passages. Therefore, they postulated that reading tests should ideally include only those comprehension items the student would get "wrong" before and "right" after reading the selections. However, Farr and Smith (1970), using this wrong–minus–right criterion for item selection on the Nelson–Denny Reading Test found it did not increase its reliability nor validity. A major problem in the wrong–minus–right criterion is that well-read readers would have relatively few items they could learn on a particular test while readers with limited informational backgrounds would have relatively many items to learn on the test. Hence, gain scores would be inversely related to pretest scores. Consequently, the search for a valid test to assess ability to learn from text has not been solved.

Another problem is that different types of measurement of comprehension seem to tap somewhat different cognitive abilities and processes. For example, use of questions stresses syntactic transformation processes; reporting information gained or retelling stories places a premium upon memory; cloze techniques emphasize ability to use syntactic and semantic constraints for inferring or predicting deleted words; paraphrasing requies verbal facility; and following directions demands not only memory but also visualization, concentration, and ability to transform verbal cues into motor activities and other forms of representation, such as iconic and schematic modes (Sticht, 1975). If different modes of assessing comprehension do tap specific abilities and processes, then readers are likely to have not only inter-individual but also intra-individual differences in reading comprehension or in learning from text. These differences are likely to be a function of students' abilities, their purposes in doing the tasks, the kinds and difficulty levels of the reading tasks, and the way in which comprehension or learning from text is assessed. Further investigation of these variables and their interrelationships is necessary.

Moreover, researchers on high school reading should specify the way in which comprehension is assessed in their studies. But, instead of limiting assessment of comprehension to paraphrasing (Anderson, 1972; Reed, 1970), investigators should simply specify the type of comprehension measure used and the conditions under which they attained their results. Then, knowledge would accumulate on the conditions under which different types of comprehension are acquired or improved.

Although problems in assessing comprehension or learning from text exist, practitioners and reading clinicians cannot wait for resolution of the problems but must instead use the best test currently available and follow statistically reliable and valid procedures for determining whether students have a reading difficulty because they are reading at a level significantly below an expectancy criterion.

Expectancy criterion. All expectancy criteria are fallible. But, in order to diagnose or evaluate students' development in reading, an expectancy criterion must of course be used. Frequently, mental age is selected for the expectancy level. This is one of the best measures of expectancy, particularly when assessed by an individual intelligence test, such as the Stanford-Binet, administered by a competent clinician. Next, the reading comprehension score converted to an age-equivalent is substracted from mental age. The discrepancy score in months is then used as an index of the degree to which the student is not reading up to potential or mental age levels. Most clinicians follow this procedure.

However, an arbitrary criterion of any number of months-discrepancy should not be used for determining those in need of reading improvement, as Ames (1969) did when he used the Lorge-Thorndike Non-Verbal Battery and the Iowa Comprehension Test to evaluate reading development. Since any expectancy criteria and any reading test have their limitations, including errors of measurement, discrepancy socres should take these measurement errors into account. Therefore, to determine whether a discrepancy score is reliable, the difference between expectancy and reading achievement should be significantly beyond chance, perhaps at least 1.96 times the standard error of difference between errors of measurement on each test. The standard should be adhered to particularly when discrepancy scores are used for placement of students in special classes.

While problems in measurement and diagnosis of reading and learning from text still remain to be solved, instructional strategies for teaching students to learn from texts have been devised. These strategies consist of advanced or cognitive organizers, adjunct questions, visual imagery, organizational and sentence patterns, contextual cues, and reading guides. We shall review the evidence on their effectiveness for improving learning from text.

Research on Instructional Strategies

Advance or cognitive organizers; structured overviews. Ausubel (1964) has claimed that organizers, obtained by previewing content at a higher level of abstraction, such as concepts and generalizations, may facilitate comprehension and retention of content at lower levels of abstraction, such as details and facts. The higher level abstractions should provide an ideational framework for subsuming the lower level content. The relationship between the advance organizer and the content it subsumes is similar to the relationship between main ideas or topical sentences stated at the beginning of a paragraph followed by sentences with supporting, illustrative, and clarifying information and examples. But, research findings from the use of advance organizers have been mixed, with only about half the studies reporting favorable effects, according to Barron and Cooper (1973). They themselves were unable to find any significant effects attributable to advance organizers in learning scientific materials. Bayuk, Proger, and Mann (1970), employing both questions and statements as advance organizers, were also unable to find significant results, perhaps because none of their organizer material appeared on the comprehension posttest, as has been the case in many other studies.

Apparently, advance organizers can have different effects upon high and low reading achievers. For example, using the term "cognitive organizer" instead of advanced organizer, Smith and Hesse (1969) observed that on a posttest containing literal and interpretive items, cognitive organizers had **negative** effects on the comprehension of good readers but **positive** effects for poor readers. Smith and Hesse postulated that good readers had their own organizational styles which conflicted with cognitive organizers, while poor readers did not have conflicting organizational strategies and, consequently, were able to benefit from the cognitive organizers.

While cognitive organizers have generally been presented in paragraph form, some researchers have experimented with an outline or diagram form called a structured overview. Barron (1969) and Earle and Barron (1973) theorized that one type of structured overview, consisting of vocabulary terms arranged in an outline or diagram, could be defined as an advance organizer because the diagram graphically revealed the logical structure among components of the subject matter. Testing this hypothesis, Estes, Mills and Barron (1969) found that structured overviews were superior to advance or organizers and to preposed questions in ninth grade biology classes, but not in ninth-grade English classes, perhaps because of differences in the nature of the content. The structured overview consisted of related biological terminology; while the advance organizer was a 500 word summary-passage more densely packed with technical terms and generalizations. The purpose questions were both specific and general. On a multiple choice biology posttest, no significant differences were found between groups receiving structured overviews and advance organizers, but the structured overview group scored significantly better than the purpose question group and the no-preparation control group. Earle (1969b) also reported favorable results for the structured overview on the understanding of mathematical relationships. On a multiple choice test of relationships between mathematical terms, significant mean differences were found between students who received structured overviews and students who did not. These results fit the educational generalization that students tend to achieve what has been taught. Therefore evaluators should at least measure what has been the specific content of a lesson.

Adjunct questions (questions inserted into text). Closely related to advance organizers are preposed questions. These test-like questions, presented **before** text passages, result in question-specific facilitation, that is, enhancement of learning and retention of answers to these specific questions only (Rothkopf, 1966). However, questions placed **after** passages produce both question-specific and general facilitation. General facilitation is defined as augmented learning and retention of both question-specific information plus additional information.

Andrews (1971, 1972–73) also found that questions placed before passages narrow the reader's focus. Furthermore, in comparison of effects of preposed questions to cognitive organizers, posttest results revealed significant superiority of cognitive organizers alone over a combination of adjunct questions and cognitive organizers.

Questions placed after passages may result in "mathematic" behaviors, that is, a more attentive and careful reading of materials following the question, even though the material is unrelated to the question (Frase, 1967, 1968a, 1969; Rothkopf, 1966; Rothkopf and Bisbicos, 1967).

Watts and Anderson (1971) suggested that if mathamagenic behavior causes a cumulative and increasing forward facilitation in textual processing, students would perform better on posttest questions over material occurring toward the end of a reading selection. However, this hypothesis was not supported by the results of,their study in which 300 high school seniors were given adjunct questions in passages on psychological principles. To account for the results, an hypothesis of **reviewing** behavior was suggested as an alternative to forward facilitation. In the same study, students given application-type adjunct questions performed significantly better on the posttest than students given simple recall questons. Watts and Anderson suggested that the students who answered application questions were more actively processing the information.

When examining frequency of adjunct question placement, Rickards and Diversta (174) found differential effects, depending upon the level of the question. Less frequent pacing (every fourth paragraph vs. every second paragraph) did not adversely affect learning at the literal level, but did have a negative influence on higher level learning.

Although adjunct questions are usually efficacious, under some conditions, they are counter-productive. For example, in analyzing results of a criterion test following reading sessions in which some subjects had adjunct questions while others did not, Strollo and Bliesmer (1973) concluded that adjunct questions may have had a negative effect on emotionally disturbed elementary and secondary students because the questions created a test-like atmosphere that was threatening to these students.

Adjunct questions also appeared to have had a negative effect on students of high ability. Shavelson, Berliner, Ravitch, and Loeding (1974) studied the effects of adjunct questions at the literal level versus questions above the literal level. They found that students with low vocabulary ability benefited more from the higher-level post-passage questions than did the students with high vocabulary ability. The investigators formulated the hypothesis that the higher level questions may act in a compensatory manner for low ability students, but interfere with high ability students' own strategies.

Adjunct questions appeared lose some of their effects when stringent restrictions of the laboratory setting are loosened. For example, in administering a criterion posttest to 58 high school seniors, half of whom read with the aid of adjunct questions, Gustafson and Toole (1970) found no significant general facilitation for adjunct questions. They hypothesized that general facilitation might have been lacking because students were given materials with adjunct questions to take home to study and, consequently, the students did not use adjunct questions as they might have done in a laboratory setting. In other words, adjunct questions may not have general facilitative effects in normal homework assignments.

The critical variable in home study have been time to read and reread the material. Testing his hypothesis, Hiller (1974) reported that students given adjunct questions scored significantly below a non-adjunct question control group on a test of incidental learning. Both groups had been given time to read and review the materials, unlike previous students where students had time to read material only once. Hiller explained that students may have studied only adjunct question-related material during review.

Investigating another influence on adjunct questions, Rothkopf and Bloom (1970) used both oral and written questions over scientific passages. On a criterion posttest, the oral question group significantly exceeded the written question group. Without considering the effects of additional semantic and syntactic information communicated by intonation for the oral group nor the interference of reading difficulties on the written group's comprehension of the questions, Rothkopf and Bloom attributed the oral group's superior performance to the effects of interpersonal interaction between students and questioner. Then, to determine if social interaction **alone** was the critical variable, Rothkopf (1972) added another group which received non-text oriented conversation instead of text-related oral questions. Because the text-related oral question group was still superior, Rothkopf rejected the **pure** social interaction hypothesis in favor of the hypothesis that facilitation occurs through orally administered **text-relevant** questions. If so, this research has negative implications for situations where teacher-student task-relevant interactions are minimized, as in programmed instruction and individual learn ing kits.

To summarize: Preposed questions tend to facilitate selection and storage of question-specific information, while postposed questions enhance retention of question-specific and general information, particularly when readers do not have time to reread or review, and do not perceive the questions as a threatening, test-like situation. Furthermore, level of thought attained tends to correspond to level of question asked. In general, adjunct questions are more beneficial to low than to high reading achievers.

Visual imagery in comprehension. Superior comprehension tends to be obtained by readers who are better able to use visual imagery in their reading processes (Anderson, 1974). Although college students who were instructed to use mental imagery performed better on a comprehension posttest than a control group (Rasco, Tennyson and Boutwell, 1975), attempts to teach students to use visual imagery have usually not been successful (Levin, 1976; Montague and Carter, 1973). Nevertheless, pictures and diagrams to accompany printed materials enhance comprehension, if they are necessary and intrinsically related to the reading passage (Bransford and Johnson, 1972), but not if they are like the usual basal reader pictures which have both relevant and irrelevant content (Samuels, 1967). Relevant pictures and diagrams may work because they depict relationships, provide an overview for organizing and assimilating content, and present abstract concepts and principles at concrete levels of thought.

Organization of passages and sentences. Organization of ideas in a passage also enhances comprehension and recall. Passages organized either by name or attributes of concepts resulted in significantly higher recall scores for subjects than did randomly ordered passages, but the most successful type of organization consisted of namely concepts followed by their attributes (Schultz and DiVesta, 1972). Another type of organization that enhanes comprehension is adapting sentence structure to oral patterns commonly used by the reader. This has been done at the fourth– (Ruddell, 1965) and the tenth-grade levels (Peltz, 1973–74). Evans (1972–73) also obtained significant comprehension scores on cloze and multiple choice tests through transformational simplication of materials read by 12th grade students.

Although students may be more successful in comprehension of simplified materials, such transformations tend to represent downward adaptation of the curriculum. Greater maturity in learning from text might result from improving students' abilities to comprehend more complext structures, but such a training study has not yet been conducted at the high school level. The best of both procedures would be initial simplification, then instruction leading toward upward adaptation of students' abilities to comprehend more complex syntactical structures. Students should not only be taught complex sentence structures, but should also be instructed in the use of intrasentence structures, incldùing use of cues within sentences.

Contextual cues. Although intrasentence cues have long been known to affect readers' comprehension (McCullough, 1972; Robinson, 1975), little is known about the relative difficulty of different kinds of cues. Attempting to find out, Dulin (1969) discovered that with a group of 10th graders, context cues could be arrayed from the easiest to the most difficult to utilize for reading comprehension in the following order: contextually-determined experiences, cause-and-effect, contrast, description, synonyms, and appositive. Further experimentation is necessary to determine how difficulty of content affect this order.

Reading guides. Another strategy for making a passage or a chapter more comprehensible is to use a reading guide. As constructed by Earle (1969a) Herber (1970b) the guides progress from literal, to intepretive, and finally to applied to levels of thought. They can be expanded to include a category for learning to read containing dictionary-type entries for new vocabulary and a category for evaluation (Singer, 1973). Expanded reading guides would consequently fit the entire progression of hierarachical cognitive and effective objectives in instruction (Barrett, 1968; Bloom, 1956; Krathwohl et al., 1964).

Reading guides are effective in helping students comprehend material that is difficult relative to students' reading abilities. For example, reading guides significantly improved students performance on a test of literary knowledge, interpretation, and application, but they were not superior to preposed questions for all students in ninth-grade social studies. However, they did significantly enhance the posttest performance of low reading achievers (Estes, 1973).

Although teacher guidance and instructional aids, such as reading guides can enhance students' abilities to learn from texts, the goal is for students to become independent of teachers, if they are to continue to learn on their own after they leave high school.

Learning How to Learn from Text

As Harlow (1949) has demonstrated, learners can learn how to learn. Indeed, teaching students to be more effective and efficient learners is a necessary goal of education, if students are to become adaptable to a rapidly changing world and knowledge-base (Buswell, 1956). Consequently, teaching students only to respond to teacher-formulated questions is not a sufficient educational objective.

Students must also be taught to formulate their own questions, and to read to answer these questions. That is, students must learn to become independent of teachers and learn from text on their own. For this purpose, teachers can use fade-in and fade-out techniques alternatively (Singer and Donlan, 1977). For example, after students have learned to respond to teacher-formulated questions at various levels of thought, these teacher-made questions can be faded-out as students learn to formulate their own questions and become more active in the learning process (Singer, 1973a). Teacher-formulated questions can then serve as a model of the kinds of questions appropriate to the purpose of instruction and the nature of the text or discipline. As students internalize the model, they can formulate their own questions and direct their own thinking and learning processes for mastering a text. One way students can improve their questioning abilities is by imitating the teacher's questions, provided, of course, teachers ask high level questions (Rosenthal, Zimmerman, and Durning, 1970).

Study skills. Study skills for self-formulation of questions and for learning from text require the student to be self-directive from the beginning. The oldest and best known study skill technique is Robinson's (1946) SQ3R technique: Survey, Question, Read, Recite, Review. This technique is well-grounded in research, not only research on SQ3R per se (Robinson, 1961), but also on each of the concepts included in it. Surveying a text or a chapter can result in perceiving its organization. This knowledge of structural relationships facilitates learning and retention (Ausubel, 1964; Katona, 1940). Self-formulated questions enable students to be active in the learning process (Singer, 1973), purposeful in reading (Glock, 1958), and focused in searching memory for information or in selecting sections of text and judging their relevancy.

Reading actively is, of course, necessary for reaching the goal of obtaining answers to one's own questions. However, whether any questions, self-formulated or teacher-posed, are helpful depends upon their relevance to test questions (Watts and Anderson, 1971). For example, Smith (1973) trained students to generate their own questions from textual aids, such as titles and subheadigs, but these students did not obtain significantly higher comprehension results than the control group. Smith speculated that the standardized test he used to assess comprehension was not sensitive enough to identify changes which may have occurred during six weeks of training.

Nevertheless, study skills can be improved, even in relatively good students. While an experimental study skills group of ninth graders improved significantly in reading over a control group, the brighter students improved the most (Smith and Burns, 1970).

Students can also become more flexible in reading rates (Fitzpatrick, 1970). Indeed, a faster rate of processing material is necessary in order for individuals slow and average in rate of reading to be able to complete lengthy high school assignments in the time available (Grob, 1970). In other words, slow and even average rates of reading can make semester reading assignments into contests in speed of reading, even though teachers may have intended to let students do the assignments at their own individual rates.

Although study skills and reading improvement courses tend to improve students' **general** reading achievement, what is also necessary is specific textual analysis and instruction in each discipline. Particularly helpful is textual analysis that requires a response (Putnam, 1974). For example, students in 20 top-level ninth-grade English classes were randomly assigned to four treatment groups. Those groups given only background information, or oral rendition of a ballad, or no instruction at all were not able to interpret the ballad as well as the group that had been told which words in the ballad to read and study carefully (Smith and Burns, 1970). Andresen (1969) also reported that his profundity scale, which takes students through a five-step literary analysis, from a physical to a philosophical plane, significantly improved literary analysis for female but not for male students, perhaps because the subject matter may have been biased towards female interests. Expanded directions, consisting of vocabulary activiteis, introduction to reading materials, and study methods, significantly aided 10th graders, working in both small and large groups, in learning biological terminology (Barron and Melnick, 1973). Using the biology textbook for its content, a study skills course enabled biology students who took the course to perform significantly better on their biology exams than the control group (Tomlinson and Tomlinson, 1975). Although another group for controlling time on the task was not used in this study, the hypothesis that specificity of training enhanced students' ability to learn from text is plausible and appealing.

Prescriptive instruction in the form of overview, processing, reaction, and reading guides is germane to the techniques developed by Herber (1970b) and his students (Herber and Barron, 1973; Herber and Sanders, 1969). Pre-underlined texts also aid learning (Grouse and Idstein, 1972), perhaps because students can then perceive which information is significant. Even though the underlining students do without instruction may help their learning, their abilities to underline or use any study technique can be enhanced. For example, Carter (1952) found that the correlation between students' and their professors' underlining of the same texts was not high.

Students not only need prescriptive techniques for masterig content but also for processes and procedures specific to the discipline encompassed by the text' indeed, instruction in reading and learning from text in content areas should also include theories, patterns of organization, modes of inquiry, concepts, technical vocabulary, devices, and symbols specific to the discipline or content area (Moore, 1969; Russell and Fea, 1963; Singer, 1970; Smith, 1965b). Thus a reading and study skills program should teach not only general but also specific contents and processes for mastering tests.

Readability research. All of the procedures listed in the above section will be limited in their effectiveness if texts are beyond students' ability levels. Apparently, most high school texts are beyond students' ability levels. Apparently, most high school texts have a reading difficulty beyond the average student's readability level. For example, Janz and Smith (1972), using the Flesch (1948, 1951) formula which is based upon sentence length, word complexity, and personal pronouns, found the majority of texts used in English, social studies, and science were too difficult for eighth, ninth, and tenth greaders in five rural high schools.

Readability data must be interpreted cautiously, particularly the cloze technique. Cloze tests are based upon the reader's ability to use contextual constraints and to retrieve vocabulary from memory for every deleted fifth word. But cloze tests do not scale passages in a linear relationship to levels of reading ability. They appear to work best when the reading passages are near the reader's difficulty, perhaps because passages that are too easy evoke more synonyms which, in orthodox cloze technique, are scored as errors. If so, then easy passages might best be scored for correct synonyms, while passages close to readers' difficulty levels can be scored for insertion of orginal word only. This curvilineary in cloze technique was reported by Smith (1972–73) who found that fourth graders read a fourth-grade passage with much better cloze comprehension than did eight to twelfth graders, but the high school student had higher cloze comprehension scores on an eight-grade passage.

Although adapting texts to students' levels of reading achievement provides for individual differences and continuity in difficulty of the reading task from grade to grade, this strategy for handling the range of individual differences can result in a detrimental, downward adaptation of the curriculum unless teachers stimulate students to develop upwards and learn to comprehend more difficult tasks. However, some high school students read so far below the level of most high school textbooks that the strategy of adapting textbooks alone will not be sufficient. Usually these students are still in the acqusition phase of reading development and in need of intensive instruction provided in special reading classes. The question is: how effective are these classes?

Effectiveness of Reading Acquisiton Classes

Classes for student reading below grade level have been in operation in high schools for over fifty years (Hill, 1971; Smith, 1965a). Schools have handled these below grades reading achievers in various ways. Some schools refer them to remedial reading clinics. Other schools may assign them to reading labs where they take tests and are placed in self-instructional programs, such as SRA Reading Labs. In other schools, remedial readers may get a crash program in a summer session course. However, the norm is for low reading achievers to be in a separate track where curriculum and texts can be adapted to their levels of reading ability, and reading instruction can be given to the total class. But special classes and the lower separate track tend to stigmatize students. This stigmatization is counterproductive to teaching and learning. Administrators must reduce or eliminate tracking, and search for ways to prevent the stigmatization attached to special classes. Nevertheless, specialized instruction is required for students in the reading acquistion phase of development at the high school level.

Two assumptions underlie reading acquisition or remedial reading programs: (a) a majority of students will make significant gains in these programs and (b) gains made in general reading ability will transfer to produce improved performance in content area courses (Evans, 1972). Although some students enrolled in such programs tend to improve very significantly, such as a two year gain after one year of instruction, the majority do not improve that much, and rate of progress upon termination of the program returns to previous levels or below, especially without continued supportive instruction. Apparently severe reading disabilities, even though ameliorated somewhat in short-term intensive treatment, are not fully correctable, regardless of staff competence in diagnosis and improvement of reading difficulties (Balow, 1965). However, significant gains do occur for some students and for these students alone current readiang acquisition or remedial reading programs continue to be justifiable (Lovell and Platt, 1963).

The assumption in evaluation of special reading programs is that students, as a consequence of these programs, will make general improvement on norm-referenced tests. However, Kelley (1973) found no significant differences on the California Reading Test after one year of instruction, even under the special conditions of performance contracting. General improvement is unlikely to occur until students improve in each specific domain sampled by the content and processes of the norm-referenced tests. Since almost all students do improve on specific criterion-referenced tests given in the program, these tests, along with standardized tests, should be used to demonstrate that specific learning has occurred whenever special classes or programs are evaluated.

Students in reading improvement classes might improve significantly on reading acquisition behavior since the task of learning to read can be mastered by almost all students, given adequate learning time. However, to determine whether improvement in general ability to learn from text has also occurred, a much longer time will be necessary for students to reduce their deficiencies in ability to gain information from text and to compete effectively with their peer group. The reason for the longer period of time and the pessimistic outlook is that learning from text consists of open-ended objectives which include such components as general information, vocabulary, and concepts. Readers who have always been at or above grade level in reading acquisition and in learning from text, and who have been reading longer and more widely, have been developing these skills and abilities in open-ended objectives throughout their school careers. No shortcut exists for overcoming this achievement gap between good and poor readers at the high school level. Consequently, intervention and, hopefully, prevention of poor reading ability as early as the primary grades is desirable. In later grades, what might perhaps help reading acquisition classes be more effective would be the use of material from content area classes in teaching acquisition skills, combined with emphasis in content area classes upon handling the wide range of individual differences. This mutual reinforcement is likely to be more effective than reading acquisition classes alone have been (Singer and Rhodes, 1977).

We have reviewed instructional strategies and theories for teaching students to learn from text and have evaluated instructional procedures at the classroom level for students in both acquisition and learning from text phases of reading development. To attempt to integrate this knowledge, we now turn to general and specific theories and models of reading.

Theories and Models of Reading

General Theories and Models

Although general theories and models of reading, including linguistic, information-processing, and developmental theories have been constructed (Singer and Ruddell, 1976), none of these theories adequately explain how students learn from texts, but some partial theories for learning from text have been formulated (Frase, 1968b; Rothkopf, 1966). What is also needed is an instructional theory for teaching students how to learn from texts. Some progress towards such an instructional theory has been made (Singer, 1971) and recently an outline of a specific, information-processing model of learning from text was constructed.

Specific Theory and Model

Miller (1975) defined reading as a problem solving task with multiple components. In addition to general processes of reading to decipher a message, Miller identified these components as being involved in learning from text: attention, retrieval of information from long term storage, and processes of elaborating and abstracting, naming, identifying, and judging relevancy of important elements of text. These components must then be integrated to form representations to the reader. This integration is aided by pictures which serve their purpose if they are relevant and contain only essential information.

Processing of text involves testing one's knowledge base for understanding by retrieval of information on a topic. Finding his knowledge base is inadequate for understanding a topic, a text-processor formulates self-questions and reads to discover new information, selecting relevant material to answer self-formulated questions. The results of these processes are integrated into verbal or ionic representations which are then stored in long-term memory.

Models such as Miller's are useful for organizing and depicting theoretically formulated and empirically determined relationships among variables, and perhaps for explaining the structures and processes involved in learning from text. Research to test and develop this model is needed. What is also needed is an instructional analogue for Miller's text information processing type of model, and research on its efficacy. We could then use it not only to integrate knowledge, but also as a practical guide for teaching students to learn from text.

Summary and Conclusion

Although high school reading has been a research topic from the beginning of research on the psychology and pedagogy of reading, and some high school reading programs have been in operation since World War I, school-wide programs, consisting of reading acquisition classes, strategies for handling individual differences in learning from text in the various content areas, and instruction for teaching students how to learn from texts on their own at the high school level, are rare. However, signs point to a dramatic change. Nine states now require high school teachers to have had a course in secondary reading. Conferences on high school and college reading are held regularly. Many texts on teaching high school reading are now available. Theories, models, and research on learning from texts at the high school level are being published with increasing frequency. All of these signs suggest that high schools are on the verge of developing school wide programs that will better provide for individual differences and will more effectively enable students to read and learn form texts at the high school level.

References

American Council on Education. *Test of General Educational Development*. Washington, D.C.: Veterans Testing Service–American Council on Education, 1970.

Ames, W. S. Reading programs: use of test data a vital factor. *Clearing House*, May 1969, *43*, 515–518.

Anderson, R. C. How to construct achievement tests to assess comprehension. *Review of Educational Research*, 1972, *42*, 145–170.

Anderson, R. C. Concretization and sentence learning. *Journal of Educational Psychology*, 1974, *66*, 179–183.

Andresen, O. The significance of profundity in literary appreciation. *Reading Research Quarterly*, 1969, *5*, 100–113.

Andrews, L. Directed reading questions and cognitive organizers: comparative effect on reading comprehension. *Research in the Teaching of English*, 1971, *5*, 79–83.

Andrews, L. Reading comprehension and three modes of prereading assistance. *Journal of Reading Behavior*, 1972–73, *5*, 237–241.

Athey, I. J. & Holmes, J. A. *Reading Success and Personality Characteristics in Junior High School Students* University of California Publications in Education, Vol. 18. Berkeley, California: University of California Press, 1969.

Ausubel, D. P. Some psychological aspects of the structure of knowledge. In S. Elam (Ed.), *Education and the Structure of Knowledge*. New York: Rand McNally, 1964, 221–262.

Balow, B. The long term effects of remedial instruction. *The Reading Teacher*, 1965, *18*, 581–586.

Balow, I. H. Does homogeneous grouping give homogeneous groups?*Elementary School Journal*, 1962, *63*, 28–32.

Balow, K. H. The effects of homogeneous grouping on seventh grade arithmetic. *The Arithmetic Teacher*, 1964, *11*, 186–191.

Barrett, T. The Barrett taxonomy: cognitive and affective dimensions of reading comprehension. In H. M. Robinson (Ed.), *Sixty-seventh Yearbook of the National Society for the Study of Education*, part 2, 1968, 19–23.

Barron, R. F. The use of vocabulary as an advance organizer. In Harold L. Herber & Peter L. Sanders (Eds.), *Research in Reading in the Content Areas: First Year Report*. Syracuse University: Reading and Language Arts Center, 1969, 29–39.

Barron, R. F., & Cooper, R. Effects of advance organizers and grade level upon information acquisition from an instructional-level general science passage. I Phil L. Nacke (Ed.), *National Reading Conference 22nd Yearbook*, 1973, 78–82.

Bayuk, R. J., Proger, B. B., & Mann, L. Organization of meaningful verbal material. *Psychology in the Schools*, 1970, *8*, 365–369.

Bloom, B. S. *Taxonomy of Educational Objectives, Handbook I:* Cognitive Domain. *New York: David McKay Company, 1956.*

Bloom, B.S. Mastery learning and its implications for curriculum development. In E. Eisner (Ed.), *Confronting Curriculum Reform*. Boston: Little, Brown & Co., 1971.

Bond, E. *Reading and Ninth Grade Achievement*. New York: Bureau of Publications, Teachers College, Columbia University, 1938.

Bond, G. L., & Tinker, M. A. *Reading Difficulties: Their Diagnosis and Correction*, Second edition. New York: Appleton–Century–Crofts, 1957.

Bormuth, J. R., Carr, J., Manning, J., & Pearson, D. Children's comprehension of between and within sentence syntactic structures. *Journal of Educational Psychology*, 1970, *61*, 349–351.

Bransford, J. D., & Johnson, M. Contextual prerequisites for understanding: some investigations of comprehension and recall. *Journal of Verbal learing and verbal Behavior*, 1972, *11*, 717–726.

Burmeister, L. E. *Reading Strategies for Secondary School Teachers*. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc., 1974.

Buros, O. K. *Seventh Mental Measurements Yearbook*. Highland Park, New Jersey: Gryphon Press, 1972.

Buswell, G. T. Fundamental reading habits: student of their development. *Supplementary Educational Monograph No. 21*. Chicago: University of Chicago Press, 1922.

Buswell, G. T. Educational theory and the psychology of learning. *Journal of Educational Psychology* 1956, *47*, 175–184.

Carter, H. D. Lecture in Educational Psychology Course, University of California, Berkeley, 1952.

Cattell, J. M. The time it takes to see and name objects. *Mind*, 1886, *11*, 63–65.

Coleman, J. The concept of equality of educational opportunity. *Havard Educational Review*, 1968, *38*, 11–22.

Cronbach, I. J. Comments on mastery learning and its implications for curricular development. In E. Eisner (Ed.), *Confronting Curriculum Reform*. Boston: Little, Brown & Co., 1971, 49–53.

Crouse, J.H., & Idstein, P. Effects of encoding cues on learning. *Journal of Educational Psychology*, 1972, *63*, 309–313.

Davis, F. *Cooperative English Test*. Princeton, New Jersey: Educational Test Service, 1960.

Dublin, K. L. M. New research on context clues. *Journal of Reading*, 1969, *13*, 33–38.

Earle, R. A. Development and using study guides. In Harold L. Herber & Peter L. Sanders (Eds.), *Research in Reading in the Content Areas: First Year Report*. Syracuse University: Reading and Lanugage Arts Center, 1969a, 71–92.

Earle, R. A. Use of the structured overview in mathematics classes. In Harold L. Herber & Peter L. Sanders (Eds.), *Research in Reading in the Content Areas: First Year Report*, Syracuse University: Reading and Language Arts Center, 1969b, 49–58.

Earle, R. & Barron, R.F. An approach for teaching vocabulary in content subjects. In Harold L. Herber & Richard F. Barron (Eds.), *Research in Reading in the Content Areas: Second Year Report*. Syracuse University: Reading and Language Arts Center, 1973, 84–100.

Estes, T. H., Mills, E. C., & Barron, R. F. Three methods of introducing students to a reading-learning task in two content subjects. In Harold L. Herber & Peter L. Sanders (Eds.), *Research in Reading in the Content Areas: First year Report*. Syracuse University: Reading and Language Arts Center, 1969, 40–48.

Estes, T. H., & Piercey, D. Secondary reading requirements: a report on the states. *Journal of Reading*, 1973, *17*, 20–24.

Evans, H. M. Remedial reading in secondary schools–still a matter of faith. *Journal of Reading*, 1972, *16*, 111–114.

Evans, R. V. The effect of transformational simplification on the reading comprehension of selected high school students. *Journal of Reading Behavior*, 1972–73, *5*, 273–281.

Farr, R. *Reading: What Can be Measured? Newark, Delaware: International* Reading Association, ERIC/CRIER, 1969.

Farr, R. & Smith, C. B. The effects of test item validity on total test reliability and validity. In G. B. Schick & M. M. May (Eds.), *National Reading Conference 19th Yearbook*, 1970, 122–134.

Fitzpatrick, E. M. *Development and Evaluation of Skimming Techniques for Teaching Flexibility in Reading to High School Students*. Doctoral dissertation, Boston University, 1970.

Flesch, R. A new readability yardstick. *Journal of Applied Psychology*, 1948, *32*, 221–233.

Flesch, R. *How to Test Readability*. New York: Harper, 1951.

Frase, L. T. Learning from prose material: length of passage, knowledge of results, and position of questions. *Journal of Educational Psychology*, 1967, *58*, 266–272.

Frase, L. T. Effect of question location, pacing, and mode on retention of prose material. *Journal of Educational Psychology*, 1968a, *59*, 244–249.

Frase, L. T. Some data concerning the mathemagenic hypothesis. *American Educational Research Journal*, 1968b, *5*, 181–189.

Frase, L. T. Structural analysis of the knowledge that results from thinking about text. *Journal of Educational Psychology Monograph*, 1969, *60* (6).

Freed, B. Secondary reading–state of the art. *Journal of Reading*, 1974, *17*, 195–201.

Gans, R. *A Study of Critical Reading Comprehension in the Intermediate Grades*. New York: Teachers College Contributions to Education, No. 811, 1940.

Gardner, E., *et. al*. Stanford Achievement Test–Advanced Paragraph Meaning. New York: Harcourt, Brace and Jovanovich, 1966.

Gates, A. I. An experimental and statistical study of reading and reading tests. *Journal of Educational Psychology*, 1921, *12*, 303–314.

Gates, A. I. & MacGinitie, W. H. *Gates-MacGinitie Reading Tests*. New York: Teachers College Press, Columbia University, 1965.

Gilbert, L. C. Functional motor efficiency of the eyes and its relation to reading. *University of California Publications in Education*, 1953, *2*, 159-232.

Gilbert, L. C. Speed of Processing visual stimuli and its relation to reading. *Journal of Educational Psychology*, 1959, *50*, 8–14.

Glock, M. Some psychological aspects of teaching reading. *Education*, 1958, *78*, 529–533.

Goodman, K. Behind the eye: what happens in reading. In H. Singer & R. Ruddell (Eds.), *Theoretical Models and Processes of Reading*, Second edition. Newark, Delaware: International Reading Association, 1976, in press.

Greene, H. A., Jorgensen, A. N., & Kelley, B. H. *Iowa Silent Reading Test: Advanced*. New York: Harcourt, Brace, & World, 1942.

Grob, J. A. Reading rate and study-time demands on secondary students. *Journal of Reading*, 1970, *13*, 285–288.

Gustafson, H. W., & Toole, D. L. Effects of adjunct questions, pretesting, and degree of student supervision on learning from an instructional text. *Journal of Experimental Education* 1970, *39*, 53–58.

Hafner, L. E. *Improving Reading in Middle and Secondary Schools*, Second edition. New York: MacMillan, 1974.

Hall, W. E. & Robinson, F. P. An analytical approach to the study of reading skills. *Journal of Educational Psychology* 1945, *36*, 429–442.

Herber, H. L. Reading in content areas: a district develops its own personnel. *Journal of Reading*, 1970a, *13*, 587–592.

Herber, H. L. *Teaching Reading in Content Areas*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970b.

Herber, H. L., & Barron, R. F. (Eds.), *Research in Reading in the Content Areas: Second Year Report*. Syracuse University: Reading and Language Arts Center, 1973.

Herber, H. L., & Sanders, P. L. (Eds.), *Research in Reading in the Content Areas: First Year Report*. Syracuse University: Reading and Language Arts Center, 1969.

Hill, W. Characteristics of secondary reading: 1940–1970. In Frank P. Greene (Ed.), *National Reading Conference 20th Yearbook*, 1971, 20–29.

Hiller, J. H. Learning from prose text: effects of readability level, inserted question difficulty, and individual differences. *Journal of Educational Psychology*, 1974, *66*, 202–211.

Holmes, J. A. Basic assumptions underlying the substrata-factor theory. *Reading Research Quarterly*, 1965, *1*, 5–28.

Holmes, J. A., & Singer, H. *The Substrata–Factor Theory: Substrata–Factor Differences Underlying Reading Ability in Known Groups*. U.S. Office of Education, final report No. 538, SAE 8176, 1961.

Holmes, J. A., & Singer, H. *Speed and Power of Reading in High School*. Cooperative Research Monograph No. 14. Washington, D. C.: U.S. Government Printing Office, Supt. of Documents Catalogue No. FS5230:30016, 1966.

Huey, E. B. *The Psychology and Pedagogy of Reading*. New York: Macmillan, 1908.

Janz, M. L., & Smith, E. H. Students' reading ability and the readability of secondary school subjects. *Elementary English*, 1972, *49*, 622–624.

Javal, E. Essai sur la physiologie de lecture. *Annales d' Oculistique*, 1879, *82*, 242–253.

Jewell, C. B. *A Reading Comprehension Test for Senior High School Students in Large Urban Areas in Alberta, Canada*. Doctoral dissertation, University of Oregon, 1969.

Judd, C. H., & Buswell, G. T. Silent reading: a study of various types. *Supplementary Educational Monograph No. 23. Chicago: University* of Chicago Press, 1922.

Karlin, R. What does research in reading reveal about reading and the high school student; *English Journal*, 1969, *58*, 386–395.

Katona, G. *Organizing and Memorizing*. New York: Columbia University Press, 1940.

Kelley, J. Effectiveness of a performance contracting program in reading and mathematics relative to educationally deprived secondary students. *Dissertation Abstracts*, 1973, p. 4567a.

Krathwohl. D. R., Bloom, B. S., & Masia, B. B. *Taxonomy of Educational Objectives: Handbook 2: Affective Domain.* New York: David McKay Company, 1964.

Levin, J. R. Comprehending what we read: an outsider looks in. In H. Singer & R. Ruddell (Eds.), *Theoretical Models and Processes of Reading*, Second edition. Newark, Delaware: International Reading Association, 1976, in press.

Lovell, K., Johnson, E., & Platts, D. A further study of the educational progress of children who had remedial instruction. *British Journal of Educational Psychology*, 1963, *33*, 3–9.

Marks, E. & Noll, G. Procedures for evaluating reading and listening comprehension tests. *Educational and Psychological Measurement*, 1967, 339–343.

Mavrogenes, N. A., Winkley, C. K., Hanson, E., & Vacca, R. T. Concise guide to standardized secondary and college reading tests. *Journal of Reading*, 1974, *18*, 12–22.

McCullough, C. M. What should the reading teacher know about language and think. In Richard E. Hodges & E. Hugh Rudorf (Eds.), *Language and Learning to Read*. New York: Houghton-Mifflin, 1972, 202–214.

Miller, G. A. (Ed.). *Semantics, Concept and Culture–Conference on Studies in Reading*. U. S. Dept. of Health, Education & Welfare. Washington, D.C.: National Institute of Education, 1975.

Monroe, M. *Children Who Cannot Read*. Chicago: University of Chicago Press, 1932.

Montague, W. E., & Carter, J. F. Vividness of imagery in recalling connected discourse. *Journal of Educational Psychology*, 1973, *64*, 72–75.

Moore, W.J. About reading in the content fileds. *English Journal*, 1969, *58*, 707-718.

Mosier, C. Batteries and profiles. In E.F. Lindquist (Ed.), *Educational Measurement*. Washington, D.C.: American Council on Education, 1951, 764-808.

Otto, W. & Smith, R.J. Junior and senior high school teachers' attitudes toward teaching reading in the content areas. In G.B. Schick & M.M. May (Eds.), *National Reading Conference 18th Yearbook*, 1969, 49-54.

Peltz, F. K. The effect upon comprehension of repatterning based on student's writing patterns. *Reading Research Quarterly*, 1973-74, *9*, 603-621.

Putnam, L.R. Don't tell them to do it . . . show them how. *Journal of Reading,* 1974, *18,* 41-43.

Rasco, R.W., Tennyson, R.D., & Boutwell, R.C. Imagery instruction and drawings in learning prose. *Journal of Educational Psychology,* 1975, *67,* 188-192.

Reed, D. A theory of language, speech and writing. In H. Singer and Robert B. Ruddell (Eds.), *Theoretical Models and Processes of Reading.* Newark, Delaware: International Reading Association, 1970.

Rickards, J.P., & DiVesta, F.J. Type and frequency of questions in processing textual material. *Journal of Educational Psychology,* 1974, *66,* 354-362.

Robinson, F.P. *Effective Study.* New York: Harper Brothers, 1946.

Robinson, F.P. Study skills for superior students in secondary school. *The Reading Teacher,* 1961, *15,* 29-33.

Robinson, F.P., & Hall, P. *Robinson-Hall Reading Test.* Columbua, Ohio: Ohio State University Press, 1949.

Robinson, H.A. *Teaching Reading and Study Strategies: The Content Areas.* Boston, Massachusetts: Allyn and Bacon, 1975.

Robinson, R.D., & Smith, B.J. Secondary consultant: remedial teacher of content teachers. *Journal of Reading,* 1973, *16,* 440- 443.

Rosenthal, T. L., Zimmerman, B. J., & Durning, K. Observationally induced changes in children's interrogative classes. *Journal of Educational Psychology,* 1970, *16,* 681-688.

Rothkopf, E.Z. Learning from written instructive materials: an exploration of the control of inspection behavior by test-like events. *American Educational Research Journal,* 1966, *3,* 241-249.

Rothkopf, E.Z. Variable adjunct question schedules, interpersonal interaction, and incidental learning from written material. *Journal of Educational Psychology,* 1972, *63,* 87-92.

Rothkopf, E.Z., & Bisbicos, E.E. Selective facilitative effects of interspersed questions on learning from written materials. *Journal of Educational Psychology,* 1967, **58,** 56-61.

Rothkopf, E.Z., & Bloom, R. D. Effects of interpersonal interaction on the instructional value of adjunct questions in learning from written material. *Journal of Educational Psychology,* 1970, **61,** 417-22.

Ruddell, R. B. The effect of the similarity of oral and written patterns of language structure on reading comprehension. *Elementary English,* 1965, **42,** 845-851.

Ruddell, R.B. Psycholinguistic implications for a system of communications model. In H. Singer & R. Ruddell (Eds.), *Theoretical Models and Processes of Reading, Second Edition.* Newark, Delaware: International Reading Association, 1976, in press.

Russell, D. & Fea, H. Research on teaching reading. In Nathan Gage (Eds.), *Handbook of Research on Teaching.* Chicago: Rand-McNally, 1963, 865-928.

Ryan, F.L., & Wheeler, R. Truth-loving mind is a sleuth-loving mind. *Instructor*, 1974, **84**, 136.

Samuels, S. J. Attentional processes in reading: the effect of pictures on the acquisition of reading responses. *Journal of Educational Psychology*, 1967, **58**, 337-342.

Schultz, C. B., & Divesta, F. J. Effects of passage organization and note taking on the selection of clustering strategies and on recall of textual materials. *Journal of Educational Psychology*, 1972, **63**, 244-252.

Shavelson, R. J., Berliner, D. C., Ravitch, M. M., & Loeding, D. Effects of position and type of question on learning from prose material: interaction of treatments with individual differences. *Journal of Educational Psychology*, 1974, **66**, 40-48.

Shores, L., & Moore, J.E. *Peabody Library Information Test.* New York: Educational Test Bureau, 1940.

Singer, H. *Substrata-Factor Reorganization Accompanying Development in Speed and Power of Reading at the Elementary School Level.* U.S. Office of Health, Education & Welfare, final report, Cooperative Research Project No. 2011, 1965.

Singer, H. Factors involved in general reading ability and reading in the content areas. In G.B. Schick & M.M. May (Eds.), *National Reading Conference 19th Yearbook*, 1970, 295-305a.

Singer, H. Theories, models and strategies for learning to read. In Fred Davis (Ed.), *The Literature of Research in Reading with Emphasis on Models.* New Brunswick, New Jersey: Rutgers University, 1971.

Singer, H. *Preparation of Reading Content Specialists for the Junior High School.* U.S. Office of Education, final report, No. OEG-0-70-1732-721, 1973.

Singer, H. IQ is and is not related to reading. In S.F. Wanat (Ed.), *Intelligence and Reading.* In preparation, 1976. Also listed in ERIC *Document Resumès*, July, 1974, ED 088 004 MF $0.75, HC $1.85.

Singer, H. Research that should make a difference. In S. Jay Samuels (Ed.), *Research for Classroom Teachers*, Newark, Delaware: International Reading Association, 1977, in preparation.

Singer, H., & Donlan, D. *Reading and Learning from Texts.* In preparation, 1977.

Singer, H., & Rhodes, A. Problems, prescriptions, and possibilities in high school reading instruction. In Wayne Otto, Charles W. Petters & Nathaniel A. Peters (Eds.), *Reading Problems in Perspective.* Reading, Mass.: Addison-Wesley, 1977, in press.

Singer, H., & Ruddell, R.B. *Theoretical Models and Processes of Reading, Second Edition.* Newark, Delaware: International Reading Association, 1976, in press.

Smith, A. E. The effectiveness of training students to generate their own questions prior to reading. In Phil L. Nacke (Ed.), *National Reading Conference 22nd Yearbook*, 1973, 71-77.

Smith N. B. *American reading instruction.* Newark, Delaware: International Reading Association, 1965a.

Smith, N. B. Reading in subject matter fields. *Educational Leadership*, 1965b, 22, 382-385.

Smith, R. J., & Burns, T. The effects of different instructional practices on student enjoyment and interpretation of a ballad. *Journal of Reading*, 1970, 13, 345-54.

Smith, R. J., & Hesse, K. D. The effects of prereading assistance on the comprehension of good and poor readers. *Research in the teaching of English*, 1969, *3*, 166-177.

Smith, W. L. The controlled instrument procedure for studying the effect of syntactic sophistication on reading: a second study. *Journal of Reading Behavior*, 1972-73, 5, 242-251.

Spache, G.D. *Test on the Use of the Dictionary*. Gainesville, Florida: University of Florida Reading Lab and Clinic, 1963.

Sticht, T. G. Theory of representation and the performance of tasks involving reading. Invitational address, National Reading Conference, St. Petersburg, Florida, December, 1975.

Strang, R. *Exploration in Reading Patterns.* Chicago: University of Chicago Press, 1942.

Strollo, J. C., & Bliesmer, E. P. Practice of effects of varying study question positions on retention of learning-handicapped students.. In Phil L. Nacke (Ed.), *National Reading Conference 22nd Yearbook*, 1973, 62-70.

Summers, E. G., Harker, J., & Trull, R. *Research and Non-research Resources in Secondary Reading.* Vancouver, Canada: University of British Colombia, 1970.

Tiegs, E.W., & Clark, W.W. *California Achievement Test – Reading.* Monterey, California: California Test Bureau-McGraw Hill, 1963.

Thorndike, E. Reading as reasoning: a study of mistakes in paragraph reading. *Journal of Educational Psychology*, 1917, 8, 323-332.

Tomlinson, B. & Tomlinson, M. Integrating Reading and Study Skills into College Biology. Paper read at the National Reading Conference, St. Petersburg, Florida, December 1975.

Watts, G.H., & Anderson, R.C. Effects of three types of inserted questions on learning from prose. *Journal of Educational Psychology*, 1971, 62, 387-394.

Wiener, M. & Cromer, W. Reading and reading difficulty: a conceptual analysis. *Harvard Educational Review*, 1967, 47, 620-643.

FOOTNOTES

[1] Reading acquisition behavior involves learning rules for responding to print and integrating these rules with syntactic and semantic processes, coupled with storage and retrieval of information and images from memory systems. Learning from text, which concurrently develops with reading acquisition behavior, is defined as initiation of behavior or changes in behavior that are relatively premanent and result from confirmation or disconfirmation of active responses to printed stimuli (Singer and Donlan, 1977).

[2] For reviews of research with emphasis on instruction in reading in content areas prior to 1970, see Herber (1970b). For a review of research in the content areas organized on a developmental basis, see Singer (1970a). Other reviews of secondary reading are also available (Karlin, 1969, Summer, Harker, and Trull, 1970).

[3] We are currently analyzing data to test this hypothesis at the University of California, Riverside.

[4] Tests are critically reviewed in Buros (1972) and Farr (1969). A guide to tests for secondary schools and colleges has also been prepared (Mavrogenes, Winkley, Hanson, and Vacca, 1974). Use of tests in research is indexed in ERIC/RCS.

SEX ROLES AND READING INSTRUCTION: A CRITIQUE OF RELATED RESEARCH

PATRICIA H. DUNCAN
Virginia Commonwealth University

Among the factors considered to influence achievement in reading, the sex of the individual has consistently received attention. Comprehensive reviews of research, such as that of Maccoby and Jacklin (1974) have shed light on actual versus assumed differences in the academic achievement of boys and girls. Johnson's (1974) study of sex differences in reading in six cultures provided a qualified confirmation of sex differences in reading achievement and created an awareness of the cultural impact upon the superiority of one sex over another as related to reading achievement.

A significant question raised by the Johnson study provides the purpose for the present review. It becomes apparent that, although difficult to separate, sex differences in achievement and perception of sex role by both student and teacher should be subjected to discrete and systematic investigation.

Following closely the advent of the women's liberation movement, the questions of sex roles and sex stereotyping have taken on increased interest. Prior to the mid-1960s research on these topics appears to be relatively sparce and scattered as reported in educational and psychological journals. It is the intent of this review to identify at least some of the investigations which might illuminate the manner in which sex role interacts with reading behavior.

Bias in Literature and Other Media

Sex stereotyping in literature has been one of the more frequently explored issues, perhaps because of the concreteness of the data. A preponderence of the research in this area has been conducted by women, which raises the question of objectivity of judgment in establishing criteria and weighing judgments according to criteria.

A survey of much public impact was reported by the affiliate of the National Organization of Women, Women on Words and Images (1972). The study, *Dick and Jane as Victims*, although simplistic in design and method, provided a comprehensive review of 134 basal readers from 15 publishers. A total of 2,760 stories were analyzed for activity mastery themes, second-sex themes, adult role models, and biographies. Among the ratios observed of male to female were the following:

Boy-centered stories to girl-centered	5:2
Adult male main character to adult female	3:1
Male biographies to female biographies	6:1

Male animal to female animal stories 2:1
Male folk or fantasy stories to
Female folk or fantasy stories 4:1

On all criteria of activity/master themes the ratios of incidence appeared higher for male characters. The study also indicated the absence in readers of themes which would provide extended dimension and flexibility to sex roles, such as sex roles across cultures and ethnic groups; excelling girls, and multidimensional boys.

In a study to compare roles, relationships, activities, and the relative importance assigned to male and female characters in readiness, first and second grade texts, Frasher and Walker (1972) analyzed 734 readers. The results of the analysis were consistent with the NOW study in that sex ratios in categories of main character, occupational roles, and activity and play were extremely disproportionate in favor of males. It was also noted that teachers manuals were found, in some cases, to perpetuate a narrow view of sex role behavior. Graebner (1972) examined 534 stories in elementary reading books published between the decade 1962–1972. Using criteria of main character, active/passive character, and boy/girl activity, she found continuing quantitative disparities in the representation of females in illustration, the number of stereotyped activities, and the number of females portrayed in passive dependent roles. She concluded that school readers are far from portraying life in light of societal changes.

Taylor (1973) reported a study of California state adopted texts used in the first grades and found 75% of the main characters to be male. Stereotyped roles were in evidence in adult characters and in activity/mastery categories. Only 15% of the illustrations included females. Taylor noted that females were seldom depicted with female peers as opposed to boys typically shown in the company of male peers. In a similar study which sought to identify racial/ethnic and sex stereotyping Britton (1975) analyzed 244 reading texts commonly used in the Northwest region and other areas of the United States. For each series and each racial/ethnic group of stories male main characters were predominant over female main characters. Percentage comparisons of sex-related career roles with minority and Anglo stories revealed that female career roles included fewer than 15% of the total.

To describe the new female protagonist in children's literature Siege (1973) critiqued 14 recent books with 20 girls in modern settings. Among some her interesting conclusions were the following:

1. The new female protagonist does not conform to the traditional heroine concept physically, intellectually, or psycholocially.
2. She is not an outstanding achiever, not delicate, frilly, not pretty, not gentle or passive/dependent, not a success in sports, studies, or social life.
3. Physical characteristics vary and differ from accepted norms, e.g.

rough, gangling, stoop-shouldered, freckle-faced, chubby.

4. Girls non conforming characterization is only a stage on their adjustment processes. Behavior becomes constructive as the environment changes for the better.

In an area considered traditionally boys' literature Tate (1973) reviewed 49 volumes of science fiction. She observed that some authors appear to be deliberately anti-femine in science fiction or avoid females altogehter. Trends in science fiction as related to sex roles were identified as: (1) female main characters in science fantasy with traditional themes; (2) social criticism with female characters demonstrating resourcefulness; and (3) brother-sister teams as main characters. Tate concluded that science fiction is ceasing to be a genre for boys only.

A study supported by the Resource Center on Sex Roles in Education (Weitzman & Rizzo, 1974) scrutinized the images of males and females in elementary school textbooks for grades one through sex. Included was an analysis of 8,000 illustrations. The following conclusions were reported as part of the study.

1. The higher the grade level, the higher the percentage of adults pictured in illustrations, reaching 73% by sixth grade

2. Whites are featured in 81% of the illustrations. The number of pictures of minority persons in texts decreases as the grade level increases.

3. Only 31% of the total textbook illustrations are of females. Again, as the grade level of the texts increases the percentage of females in illustrations decreases.

4. Image stereotypes included active, adventurous, independent, emotionally constrained boys and passive, dependent, narcissistic, emotional girls.

It would appear that sex bias in children's literature and in reading texts persists; however it seems as though women themselves must assume much of the responsibility for this since they constitute a majority of the authors (Rinsky, 1973). Publishing houses have recently taken major steps to eliminate sex role stereotyping in written material, but some sex biases are so subtle that they escape even the most prudent of publisher control. It then becomes the job of the teacher to be sensitive to bias in the selection and use of material. One would hope that Heilbrun's (1973) caution would be heeded that stories be selected with strong themes, plots, and characters as priority criteria.

Aside from the influence of written narrative, television has been observed to play an important part in influencing the development of sex roles in children. Sternglanz and Serbin (1974) conducted an observational analysis of the sex role models show on ten children's television shows. Characters were categorized according to major/minor roles, sex, goodness/badness, and importance in the plot. Significant differences were found both in the number of male and female roles (favoring males) and in the behaviors emitted by male and female characters. Males were shown to be agressive and constructive while females were shown to be dependent. Males were rewarded more often for activity. Magic was often presented as a manipulative device for girls. In a follow-up study to the preceding study Fruch and McGhee (1974) investigated the extent to which the amount of time watching television was related to the strength of traditional sex role development in kindergarten, second, fourth, and sixth grade children. Amount of viewing time was compared with scores on the IT scale for sex role preferences. Children with higher television watching time scored significantly higher on the IT scale. Lack of interaction effects suggested that this relationship holds equally for boys and girls and does not change with increasing age.

Sex Role Perception as it Affects Achievement

That children acquire sex-stereotyped behavior at an early age seems to be fairly well established. Lewis (1972) found that boy and girl infants by the age of six months have been differentially reinforced for response to auditory stimuli, proximal/distal behaviors (which may form the basis for later language learning), and social behaviors. Keogh (1972) in an investigation of spatial orientation on drawing and walking tasks speculated that difference in task orientation support the contention that sex difference in perceptual and cognitive styles may be identifiable by the late preschool period. This hypothesis was not supported by Turnure (1975) in a study of seven, nine, and twelve year olds as measured by social role taking and Piagetan tasks. Differences by sex were not statistically significant on either types of tasks.

Other studies of preschool sex-role behavior have indicated the resistance of boys to sex-inappropriate behavior (Ross & Ross, 1972). To assess capacity for identifying sex stereotypes Nadelman (1974) asked 240 kindergarten and third grade children to respond to tasks of verbal recall, knowledge, and preference. On each task children were able to identify more same-sex items with more rigid behavior occuring when choices were limited. Middle class children performed better on the knowledge test than did working class children.

Among the early influences on reading achievement vocabulary and language usage rank high. Bias in vocabulary and oral expression are extremely subtle and resistant to modification. In the revision of their *Adjective Check List* (which is a measure of knowledge of sex stereotypes) Williams, Bennetts, and Best (1975) administered the instrument to 284 Euro-American kindergarten, second, and fourth grade children. They found that kindergarten children show an appreciable degree of knowledge of adult sex stereotypes, and that this knowledge increases rapidly until second grade, followed by a "latency period." The authors suggested that the stereotyping in basal readers may not have as much influence as expected in that knowledge of sex stereotypes seems to be well established at preschool and primary ages. Larrick and Merriam (1973), in referring to their new anthology, *Male and Female Under 18*, found among their 3,000 responses a continuing tendency for elementary age girls to accept the stereotype roles for females, and for elementary age boys to confirm the image of Girls/Trivial–Boys/Significant. Tiedt and Semorile (1973) reviewed the semantic descrepancies in language with terms becoming sex-specific by association. Gunderson (1972) warned that, "Children interpret language quite literally, and when they hear expressions such as **chairman**, or **bortherly love** they think of men rather than of the human race." A brief study by Kramer (1974) showed that the folk hypothesis that women differ from men in use of language was not in evidence in the writing of 34 college students.

These students from related disciplines suggest that sex-role learning occurs early in areas which may affect reading, specifically in the development of perceptual/cognitive styles, in the choice of materials, and in the acquisition of fixed concepts of vocabulary. As sex roles relate more directly to reading. Gunderson listed the following factors as contributing to the reading difficulty of boys: (1) predominance of females as teachers in the primary grades; (2) boys lack of interest in basal readers; (3) adults considering reading to be a female activity; (4) effect of the differing socialization processes of males and females upon school success; and (5) teacher attitude toward boys during reading instruction. In a study which confirms a substantial boy of research on the positive correlation of auditory-visual integration with early reading achievement, Reilly (1971) found that girls develop these perceptual abilities earlier and to a higher level of success than do boys. This is worth noting in light of the study cited earlier (Lewis, 1972) which observed the selective reinforcement of baby girls' responses to auditory stimuli by mothers.

An earlier study by Mazurkiewicz (1960) dealt directly with the relationship between sex roles and academic achievement. Using eleventh grade boys, he found that reading achievement test scores were higher for those boys who considered reading a masculine activity. In a study of the relationship of attainment values, achievement behavior, and success expectation, Stein, Pohly, and Mueller (1971) established that sex appropriate tasks influenced boys' behavior more than girls' at the sixth grade level. Girls with high preference for masculine roles devoted more time and effort to masculine rather than feminine tasks.

Dwyer (1974) sought to determine the influence of children's sex role standards on reading and arithmetic achievement. She compared the scores of 385 students (grades 2, 4, 6, 8, 10, and 12) on two measures of sex role preference and on the *Stanford Achievement Test* and *ITED* subtest scores in reading and arithmetic. The effect of one measure of sex role standards (SRS) accounted for 4.5% of the variance in reading achievement when considered in combination with other variables. With other variables partialed out it accounted for 1.5% of the total variance, a figure significant at the .001 level. The effect of sex role standards was stronger for males than females. The effect of standards about reading on reading achievement was observed at all grade levels tested and was found to be strongest at grade levels six and above. Among the interesting conclusions of this study were some of the following:

(a) Subjects considered reading more feminine than masculine.
(b) Stronger social sanctions exist against males who participate in female roles.
(c) Boys were more likely to label things exclusively male, but the reverse was not true.
(d) Younger subjects showed more rigidity concerning sex roles than did older subjects.

Perhaps as the child develops with age a greater security in his/her sexual identity, he/she is able to deal more flexibly with the issue in relation to academic tasks.

A comprehensive survey of studies of children's reading preferences by Tibbetts (1974) led her to conclude that social pressure may be the factor behind children's reading choices instead of sex differences. Ego rather than sex was determined to be the most important factor governing choice and there was an indication that there is a tendency for both sexes to avoid identification with an inferior sex role.

Sex of Teacher as a Model

The impact of the teacher model in the reading instructional setting should be weighed carefully. Wolf (1973) emphasized the support that a same-sex model lends to a social learning interpretation of sex typing. Friedman and Bowers (1971) noted that a child will continue to imitate a model's behavior on occasions when direct reinforcement is not provided. To demonstrate the power of the teacher-model Friedman and Bowers (1971) observed 120 urban preschool, kindergarten, and first grade childen to determine the extent to which the children imitated the teacher's verbal behavior. First grade children showed significantly more imitative behavior than did the younger children. Significant sex differences showing greater imitation by girls occurred in four categories of verbal behavior and increased with grade level.

To assess the characteristics of model imitation in first grade children Garrett and Cunningham (1974) asked 48 boys and 48 girls to select pictures after observing the selections of pictures made by a model. Relevant to sex role were the findings that (a) there was no tendency for either sex to be a more salient model or experimenter; (b) Same-sex models plus reward condition produced the greatest imitation; (c) Negative conditions appear to be perceived with model/experimenter combinations which resemble the traditional parent relation stereotype of an authoritarian father and a passive mother.

A significant study by Asher and Gottman (1973) explored the effect of sex of teacher on the reading achievement of fifth graders. Pretest and posttest data were compared for 534 students in 34 elementary schools. Results indicated the familiar pattern of superior reading among females. Instruction by a male or female teacher did not affect the reading performance of boys and girls. A replication of this study during the following year confirmed the stability of the results of the initial study. The authors suggested that further research on the classroom environment should be investigated rather than assume that achievement variables are correlated with the sex of the teacher.

Samuels and Turnure (1974) studied the relationship of attending behavior to the reading achievement of first grade children as measured by a sight word test. Observers recorded attentional behavior during the reading periods. It was found that girls were significantly superior to boys in word recognition and in classroom attentiveness. Samuels and Turnure cited studies of cultural differences in scholastic achievement to account for the fact the American female teachers may influence the attentive behavior of girls.

In an effort to determine the effect of male reading tutors on second, third, and fourth grade Mexican-American boys, Gentile (1975) randomly assigned 60 of the boys to male and female tutors (half Anglo and half Mexican-American). Pretest-posttest comparisons on the *Gates MacGinitie Tests* and the *Classroom Reading Inventory* (CRI) revealed a differential effect for tutor sex at the second and third grade levels. For the second grade group the female group showed greatest gains, while the male-tutored group performed better at the third grade. Fourth Graders made greater gains in all groups. The study is inconclusive as to the effect of male tutors with Mexican-American boys who were assumed to manifest strong masculinity views.

Sex Role as it Affects Teacher Expectancy for Student Achievement.

Levitin and Chananie (1972) hypothesized that 40 white primary teachers would respond most positively to the sterotyped agressive-boy and dependent-girl models. Each teacher was asked to rate two hypothetical children performing one of three behaviors – dependency, agression, achievement. The response to sex-typed behavior indicated that the teachers responded most favorably to the achievement variable regardless of the sex of the child. In only one instance did the teachers indicate a sex-related response trend, and that was in preference for dependent versus agressive girls. A subsequent study by Etaugh and Hughes (1975) examined the roled of teacher sex and school setting in determining teacher response to hypothetical portrayals of sex-typed behaviors. The subjects were 64 female and 64 male teachers of grades five through eight. Replicating the methods and procedures of Levitin and Chananie (1972), teachers rated the children on seven-point scales indicating (a) approval of the behavior, (b) liked the child, (c) considered the behavior typical for the age and sex of the child. Both male and female teachers showed approval of dependency over agression for both boys and girls. Male teachers showed even greater approval of dependency than did female teachers, but liked children in all categories more than female teachers did.

Exploring race and sex as concomitants of a composite halo in teacher rating of pupils, Jaeger and Freijo (1975) tabulated the ratings of all teachers (22,067) in 2,920 schools for 104,036 students for students in grades two, four, and six. In contrast to the hypothesized ranking, white females rating blacks showed the least composite halo and blacks rating blacks showed the most. It was suggested that perhaps those who value teaching as a career place a high priority on the social worth of others. It was also suggested that blacks reaching middle class status often perceive themselves to be socially distant from others. Black female teachers showed relatively little composite halo in rating white students. Although this study emphasized racial aspects of the "halo" effect in teacher evaluation, it did point to the fact that the social perceptions of teachers cannot be classified uniformly with other social groups.

Miller and Herring (1975) attempted to answer the question of the relationship of student placement to teacher preference for teaching a high or low reading group. They further investigated the preference of first grade teachers to teach boys of girls. Results demonstrated that the 26 first grade teachers preferred to teach the better readers. There was no support for the belief that teachers prefer to teach one sex over another. It appears from these few studies that teacher value and reward achievement behavior over traditional sex-role behavior, and that differential expectancies by sex are similar for male and female teachers. That teachers interpret student behavior in sex-stereotyped ways seems evident; however, matching the sex of the student and teacher has not yielded significant changed in role perception. Perhaps the absence of sufficient numbers of male teachers at some age levels prevents accurate comparisons with regard to teacher expectation at this time.

References

Asher, S. & Gottman, J. Sex of teacher and student reading achievement. *Journal of Educational Psychology*, 1973, 65, 168-171.

Bernstein, J. The changing roles of females in books for young children. *The Reading Teacher*, 1974, 27, 545-549.

Britton, G. Danger: state adopted texts may be hazardous to our future. *The Reading Teacher*, 1975, 29, 52-58.

Brophy, J. & Good, T. Teacher's communication of differential expectations for children's classroom performance: some behavioral data. *Journal of Educational Psychology*, 1970, 61, 365-374.

Dwyer, C. Influence of children's sex role standards on reading and arithmetic achievement. *Journal of Educational Psychology*, 1974, 66, 811-816.

Etaugh, C. & Hughes, V. Teacher's evaluations of sex-typed behaviors in children: the role of teacher sex and school setting. *Developmental Psychology*, 1975, 11, 394-395.

Frasher, R. & Walker, A. Sex roles in early reading textbooks. *The Reading Teacher*, 1972, 25, 741-749.

Friedman, P. & Bowers, N. Student imitation of a rewarding teacher's verbal style as a function of sex and grade level. *Journal of Educational Psychology*, 1971, 62, 487-491.

Fruch, T. & McGhee, P. Traditional sex role development and amount of time spent watching television. *Developmental Psychology*, 1975, 11, 109.

Garrett, C. & Cunningham, D. Effects of vicarious consequences and model and experimenter sex on imitative behavior in first grade children. *Journal of Educational Psychology*, 1974, 66, 940-947.

Gentile, L. Effect of tutor sex on learning to read. *The Reading Teacher*, 1975, 28, 726-730.

Good, T. & Brophy, J. Behavioral expression of teacher attitudes. *Journal of Educational Psychology*, 1972, 63, 617-624.

Good, T., Sikes, J. & Brophy, J. Effects of teacher sex and student sex on classroom interaction. *Journal of Educational Psychology*, 1973, 65, 74-87.

Graebner, D. A decade of sexism in readers. *The Reading Teacher*, 1972, 26, 58.

Gunderson, D. Sex roles in reading. Paper Presented at the Annual Meeting of the American Educational Research Association, April, 1972, ED 064 671 EDRS, 10 pp.

Heilbrun, C.C. Human women. *The New York Times Book Review*, July 15, 1973, pp. 8, 10, 12.

Jaeger R. & Freijo T. Race and sex as concomitants of composite halo in teacher's evaluative rating of pupils. *Journal of Educational Psychology*, 1975, 67, 226-237.

Johnson, D. Sex differences in reading across cultures. *Reading Research Quarterly*, 1973-4, 9, 67-86.

Keogh, B. Spatial Organization of young children. ED 063 042, 42pp.

Kramer, C. Folk linguistics: wishy washy mommy talk. *Psychology Today*, June 1974, 82-85.

Larrick, N. & Merriam, E. Open letter to teacher of girls. *Elementary English*, 1973, 50, 1021-1023.

Levitin, T. & Chananie, J. Responses of female primary school teachers to sex-typed behaviors in male and female children. *Child Development*, 1972, 43, 1309-1316.

Lewis, M. Sex stereotypic behavior in infants: an analysis of social-interpersonal relationships. ED 057 904, 17 pp.

Maccoby E. & Jacklin, C. Myth, reality and shades of gray: what we know and don't know about sex differences. *Psychology Today*, 1974, 8, 109-112.

Mazurkiewicz, A. Social-cultural influences and reading. *Journal of Developmental Reading*, 1960, 3, 254-263.

Miller, H. & Hering, S. Teacher's ratings – which reading group is number one? *The Reading Teacher*, 1975, 28, 389-391.

Nadelman, L. Sex identity in American children: memory, knowledge, and preference tests. *Developmental Psychology*, 1974, 10, 413-417.

Reilly, D. Auditory-visual integration, sex, and reading achievement. *Journal of Educational Psychology*, 1971, 62, 482-486.

Rinsky, L. Equality of the sexes and children's literature. *Elementary English*, 1973, 50, 1075 & 1092.

Ross, D. & Ross, S. Resistance by preschool boys to sex-inappropriate behavior. *Journal of Educational Psychology*, 1972, 63, 342-346.

Samuels, J. & Turnure, J. Attention and reading achievement in first-grade boys and girls. *Journal of Educational Psychology*, 1974, 66, 29-32.

Siege, S. Heroines in recent children's fiction – an analysis. *Elementary English*, 1973, 50, 1039-1043.

Stein, A., Pohly, S. & Mueller, E. The influence of masculine, feminine, and neutral tasks on children's achievement behavior. *Child Development*, 1971, 42, 195-207.

Sternglanz, S. & Serbin, L. Sex role stereotyping in children's television programs. *Developmental Psychology*, 1974, 10, 710-715.

Tate, J. Sexual bias in science fiction for children. *Elementary English*, 1973, 50, 1061-1064.

Taylor, M. Sex-role stereotypes in children's readers. *Elementary English*, 1973, 50, 1045-1047.

Tibbetts, S. Sex differences in children's reading preferences. *The Reading Teacher*, 1974, 28, 279-281.

Tiedt, P. & Semorile, R. Semantic Discrepancies. *Elementary English*, 1973, 50, 1065-1066 & 1074.

Turnure, C. Cognitive development and role-taking ability in boys and girls from 7 to 12. *Developmental Psychology*, 1975, 11, 202-209.

Weitzman, L. & Rizzo D. Biased textbooks: images of males and females in elementary text books in five subject areas. Report of The Resource Center on Sex Roles in Education, 1974, 11pp.

Williams, J., Bennett, & Best, D. Awareness and expression of sex stereotypes in young children. *Developmental Psychology*, 1975, 11, 635-642.

Wolf, T. Effects of live modeled sex-inappropriate play behavior in a naturalistic setting. *Developmental Psychology*, 1973, 9, 120-123.

Women on Words & Images. *Dick and Jane as Victims: Sex Stereotyping in Children's Readers*, Princeton, New Jersey: Carolingian Press, 1972, 57pp.

SYNTAX AND READING COMPREHENSION DO THEY HAVE A FUTURE TOGETHER? ? ?

PATRICIA M. CUNNINGHAM
Almance County Schools

For over a decade now, psychologists, linguists and reading researchers have conducted investigations designed to clarify the term "syntax" and to illuminate its assumed relationship to facility with spoken and written language. Numerous studies have been carried out in which subjects performed such varied tasks as reading syntactically violated sentences (Siler, 1973-74), reporting where they heard clicks in taped prose selections (Scholes, 1969) and reading stories presented in different modes (sentences, words, phrases, fragmented groupings) (Cromer, 1970). The results of most of these experiments support the common-sense notion that there is a relationship between syntactic structure and comprehension. The nature of this relationship, however, is little clarified by the multitude of empirical investigations or the summaries of these investigations. This lack of clarification is, in part, due to lack of focus. The purpose of this paper is to focus on two specific questions: (a) What is the relationship between frequency of occurence of syntactic structures in speech and writing and reading comprehension? and (b) Are there specific syntactic structures which appear to impede reading comprehension? In the remainder of the paper, research which directly addresses one of these two questions will be critically reviewed and a tentative answer to the focus questions will be suggested.

Strickland (1962) observed that children's oral language development was more advanced than the sentence patterns being used in beginning readers and suggested that children might have difficulty reading the shorter, less familiar language of the basal. Ruddell (1965) investigated the effect of the similarity of oral and written patterns of language structure on reading comprehension. He selected high and low frequency patterns from the Strickland study and constructed passages characterized by these patterns. As measured by a cloze test, fourth-graders' comprehension of high frequency passages was significantly greater than their comprehension of passages containing infrequent patterns.

The validity of this study was questioned by Clymer (1965) who reminded the reader that cloze is more a measure of redundancy of ideas than of reading comprehension. Results obtained by Tatham (1968), however, confirmed Ruddell's findings. Fifty-three fourth graders read two passages comparable in length, readability, content and style and answered multiple choice comprehension questions. The comprehension of the passage containing the high frequency patterns was significantly greater than comprehension of the passage written with low frequency patterns. Tatham (1970) investigated the same question using second and fourth graders and a picture matching task to measure comprehension. The superiority of comprehension of the high frequency passage was again confirmed, this time at the .001 level of significance.

Sauer (1970) used nonsense language to determine which grammatical structures fourth graders understood best. Her sample of 153 midwestern lower-middle class subjects was asked to translate sentences such as "A boing boinged our boing" into English. Sauer concluded that the easiest sentence patterns were those containing a noun and verb or a noun, verb and object. These were high frequency patterns according to the Strickland study. Sentences containing a noun, verb, object and either an indirect object or a compliment were more difficult to translate.

Smith (1971) used the data from Hunt's (1966) study to write passages in the style of "typical" 4th, 8th and 12th graders. The vocabulary, content and sentence length of these paragraphs were held constant; only the level of syntactic maturity was varied. Cloze passages of these paragraphs were administered to students in grades 4-12. The results of this experiment revealed that while older students read all levels of writing significantly better than did the younger students, 4th graders read the 4th-grade writing best and 11th graders read the 4th-grade writing with least facility. Smith suggests that students may read with most facility material written at a syntactic level similiar to the one they are accustomed to reading and writing.

Reid (1972) investigated the relationship between familiar syntactic structures and sentence comprehension with seven- and eight-year-old Scottish children. Two forms of a test were constructed in which a sentence was followed by two questions. Form A of the test consisted of 19 sentences taken from a school reader. On Form B, these sentences were rewritten in a form described by Reid as closer to the spoken language of the children. A comparison of the medians on the two tests favored the responses to the rewritten questions at the .0005 level.

Peltz (1974) tested the effect of repatterning passages on the reading comprehension of 10th grade students. Passages from a 10th grade social studies texts were repatterned to approximate the syntactic patterns found in a transformational analysis of the writing of the 10th grade subjects who would read the text. The subjects, 34 randomly chosen average readers, completed eight cloze and four multiple choice tests. Each subject alternated between unaltered and repatterned passages, completing four cloze and two multiple choice tests on each. The results of the one-way analyses of variance indicated that when measured by cloze, the repatterned content material had a significant

(.05 level) effect on comprehension. When measured by multiple choice tests, however, there were no significant differences between test scores on the original passages and test scores on the restructured passages. Peltz suggested that these results may have been a consequence of the investigator's having constructed too difficult a multiple choice test instrument or that it may suggest that multiple choice and cloze tests are measuring different comprehension factors.

The answer to the question of the relationship between frequency of occurence of syntactic structures in oral and written language and reading comprehension seems to be that there is one. The results of the investigations of Ruddell, Tatham, Sauer, Smith, Reid and Peltz all support one another and strongly suggest a positive relationship between the sentence patterns used in the expressive functions of speech and writing and the receptive act of comprehending written material. This relationship seems to exist for children from primary school through high school and in at least one investigation in a foreign country.

While the existence of this relationship may have many implications for reading instruction, two practices seem particularly affected by it. The first is vindicated; the second, questioned. The vindicated practice is the use of the Language Experience Approach in beginning reading instruction. This approach, advocated by Stauffer (1969) and Allen (1964) among other, builds on the oral language of the child. A tenet of this instructional approach is that the language used by the child is written down exactly as he says it. There can be no mismatch between the syntax of the reader and that of the reading material if this method of beginning reading instructions is followed.

The questionable practice is the use of readability formulas which use word difficulty and sentence length (Dale and Chall (1948); Fry, (1968), for example) to determine the average reading level of a book. Both the finding that material is more easily comprehended if its syntax is similiar to the child's spoken and written syntax and the related finding that some syntactic structures seem to be inherently difficult suggest the need for including a measure of syntactic complexity in readability formulas.

The second questions on which this paper will now focus its attention is the identification of specific syntactic structures which impede reading comprehension. Several investigators have attempted to isolate structure causing difficulty at particular grade levels.

Fagan (1971) analyzed the effects of 43 transformations on the reading comprehension of fourth, fifth and sixth graders. Reading comprehension was measured by a cloze test. His findings indicate that the presence of deletion and embedding transformations tended to make sentences and passages difficult for children to read. Specific transformations causing the most difficulty were appositives, ing-nominalizations, genitive pronouns, common-elements deletions and negatives. Fagan also found that the number of transformations within a sentence did not seem to affect the comprehension difficulty of that sentence and that due to redundancy, sentence difficulty was more dependent on the presence and difficulty of transformations than was passage difficulty.

Marcus (1969) developed a diagnostic test of sentence meaning and administered it to 421 fifth, sixth, seventh and eighth graders who scored 5.0 or above on a word recognition test. Six items were included for each of 17 grammatical structures. He concluded that the average number of correct items increased as grade level increased and that a difficulty order could be established. Structures receiving incorrect responses from more than 35% of the sample were: Prepositional Phrase Modifiers, Complex sentences with 2 Relative Clauses, Active Verbs Transformed into Nominalizations, Relative Clauses Modifying Subjects or Objects of Prepositions, and Subjective Complements Embedded as Modifiers.

Bormuth, Carr, Manning, and Pearson (1970) investigated fourth graders' reading comprehension of between- and within-sentence syntactic structures. In this study, a comprehension skill was defined as the ability to respond correctly to a question beginning with the letters "wh" which deletes one of the constituents of a syntactic structure. Written responses to "wh" questions were gathered from 240 semirural midwestern fourth graders. Structures investigated included 25 types of sentence structures, 14 Anaphoric structures (pronoun-like structures, comprehension depends on knowing referent of pronoun) and 16 itersentence structures (relative sequential position of sentences and conjunctive expressions signal information about the relationships of the sentences to each other). The intersentence category of questions was significantly more difficult than both the sentence questions and anaphoric questions. Sentence structures which caused difficulty for 25% or more of the students were: Comparitive, equal (as), Subordinate sentences (as, if, before), Adjective compliments, Relative clauses (without deletion and appositives), Nominalizations (for-to) and certain Nominal compounds. Anaphoric structures causing comprehension difficulties for more than 25% of the sample were: Personal pronouns, Negated pronouns, Demonstrative clauses and Semantic Substitutes. All but one of the inter-sentence structures caused difficulty for 25% or more of the sample.

Bormuth et al. (1970) concluded that "by far the most startling result of this study was the fact that large proportions of the students were unable to demonstrate a comprehension of the most basic structures by which information is signaled in language" (p. 355). They further conjectured that any efforts to generalize these results to the general school population suggest an even grimmer picture since their sample did not represent the population typically consisted in serious trouble with respect to another achievement.

Robertson (1968) investigated pupils' understanding of connectives in reading. A stratified random sample of 402 fourth, fifth and sixth graders was administered the Connectives Reading Test, a test of mental ability and a standardized achievement test. The Connectives Reading Test consisted of 150 sentences with multiple choice endings designed to test pupils' understanding of 17 different connectives. Results indicated that pupils' understanding of connectives was significantly related to sex; mental age; listening, reading and writing ability; and increased significantly with an increase in grade level. Six connectives which caused difficulty for 34% or more of the students were: **however, thus, although, which, and** and **yet.**

Stoodt (1970) attempted to determine if an understanding of conjunctions was related to reading comprehension. A stratified random sample of fourth graders was administered a multiple choice test of conjunctions and three cloze passages. One cloze passage had a large number of conjunctions; the second had only ½ as many and the third contained no conjunctions. Standardized reading achievement and mental maturity tests were also given to the subjects. A significant relationship between reading comprehension and understanding conjunctions was found for both the multiple choice and the cloze measures. The most difficult conjunctions were: **when, so, but, or, where, while, how, that** and **if.**

Since the research on the reading comprehension difficulty of specific syntactic structures has been very limited, the following conclusions are most tentative. Relative clauses appear to obscure meaning especially those containing appositives and "wh" words. Nominalizations and complements are sources of comprehension difficulty. Subordinate clause linked by connectives to the main cluase may cause difficulty especially when expressing time or causality. As mentioned previously, constructors of readability formulas should attempt to include syntactic factors in their readability estimates. Teachers of reading should be alerted to these structures and help students to paraphrase passages in which comprehension difficulties seem to relate to specific syntactic structures.

References

Allen, R. V. The Language Experience Approach. In W. G. Cutts (Ed.) *Teaching Young Children to Read*. Washington, D.C.: U. S. Office of Education, 1964.

Bormuth, J. R., Carr, J., Manning, J., & Pearson, D. Children's comprehension of between- and within-sentences syntactic structures. *Journal of Educational Psychology*, 1970, **61**, 349-357.

Clymer, T. Comments. *The Reading Teacher*, 1965, **18**, 275.

Cromer, W. The difference model: A new explanation for some reading difficulties. *Journal of Education Psychology*, 1970, **61**, 471-483.

Dale, E., & Chall, J. A formula for predicting readability. *Educational Research Bulletin*, 1948, **27**, 11–20; 37–54.

Fagan, W. T. Transformations and comprehension. *The Reading Teacher*, 1971, **23**, 169–172.

Fry, E. B. A readability formula that saves time. *Journal of Reading*, 1968, **11**, 575–578.

Hunt, K. W. Recent measures of syntactic development. *Elementary English*, 1966, **43**, 732–739.

Marcus, A. D. The development of a diagnostic test of syntactic meaning clues in reading. Doctoral dissertation, New York University. Ann Arbor, Michigan: University Microfilms, 1969.

Peltz, F. K. The effect upon comprehension of repatterning based on student's writing pa tterns. *Reading Research Quarterly*, 1973–74, **9**, 603–620.

Reid, J. F. Children's comprehension of syntactic features found in some extension readers. In J. F. Reid (Ed.), *Reading: problems and practices*. London: Ward Lock, 1972, 394–403.

Robertson, J. E. Pupil understanding of connectives in reading. *Reading Research Quarterly*, 1968, **3**, 387–416.

Ruddell, R. B. The effect of oral and written patterns of language structure on reading comprehension. *The Reading Teacher*, 1965, **18**, 270–274.

Sauer, L. E. Fourth grade children's knowledge of grammatical structures. *Elementary English*, 1970, **47**, 807–813.

Scholes, R. J. Click location judgments. *Communication Sciences Laboratory Quarterly Report*, 1969, **7** (1), 33–38.

Siler, E. R. The effects of syntactic and semantic constraints on the oral reading performance of second and fourth graders. *Reading Research Quarterly*, 1973–74, **9**, (4), 583-602.

Smith, W. L. The effect of transformed syntactic structures on reading. In Braun (Ed.), *Language, Reading and the Communication Process*. Newark: International Reading Association, 1971. 52–62.

Stauffer, R. G. *Teaching Reading as a Thinking Process*. New York: Harper and Row, 1969.

Stoodt, Barbara D. *The relationship between understanding grammatical conjunctions and reading comprehension*. Unpublished doctoral dissertation, Ohio State University, 1970.

Strickland, R. G. *The Language of Elementary School Children: Its Relationship to the Language of Reading Textbooks and the Quality of Reading of Selected Children*. Bulletin of the School of Education, Indiana University, Bloomington, Indiana, 1962, 38.

Tatham, Susan M. Reading comprehension of materials written with select oral language patterns: A study at grades two and four. *Reading Research Quarterly*, 1970, 5, 402–412.

Tatham, Susan M. Using multiple choice questions to measure the effect on comprehension of materials written with select oral language patterns 1968. (ERIC Document Reproduction Service No. ED 018 355)

THE AGING PROCESS AND ITS RELATIONSHIP TO READING: A REVIEW OF LITERATURE FROM GERONTOLOGY WITH IMPLICATIONS FOR FUTURE RESEARCH

RICHARD D. ROBINSON
University of Missouri-Columbia

GERALD MARING
University of Missouri-Columbia

Within recent years there has been an increased interest among researchers from various fields in the study of the aging process. While much of this research has been primarily concerned with geriatrics there is a growing number of studies investigating other areas of gerontology. Included in this work have been a number of papers addressing themselves to the aging process and its implications for reading. Because most of this research has been reported in non-traditional reading sources, it was the authors purpose to bring together in this paper those references considered to be the most relevant. The following criteria were used in selection: (1) importance of the study in relation to the total field, (2) extent and usefulness of the bibliography, and (3) implications of the study for future research.

Definition of the Aged

A major unresolved problem associated with the study of the aging is a comprehensive definition of the population under consideration. Some would describe this group strictly according to age, others would emphasize physical norms, and still others job and social situation. Most of the research reviewed for this study generally followed the definition most commonly used by the Department of Health, Education, and Welfare which simply includes any person 65 years or older.

Number of Older Americans

The most recent census figures, as reported by *Facts About Older Americans* (1975), indicated that in 1974 one in every 10 Americans was 65+, that is, 21.8 million men and women. In addition, at present death rates, this age group is anticipated to increase 40% to 31 million by the year 2000. Were the present low birth rate to continue in the near future, these 31 million persons would be 11.7% of the total population, estimated to be about 262 millions.

General References Including Bibliographies, Surveys, and Journals

Bibliographies

Numerous bibliographies have been compiled on various aspects of the aging process. For those seeking an overview of the current literature on aging some of the most useful collections are the ones developed by the U.S. Department of Health, Education, and Welfare. *Words on Aging – A Bibliography* (1970) is the seventh edition of selected annotated references on aging published since 1950. This edition is particularly important because it includes relevant periodical articles from the previous six issues covering the period from 1963 through 1967 and selected books from 1900 through 1967. *More Words on Aging-Supplement* (1971) reviews books and periodical articles during the period 1968 through 1970.

Several bibliographies have been concerned primarily with research and doctoral study and the aging. *Education For Aging* (1975) contains 100 abstracts of research documents related to various aspects of the aged. This reference should be especially useful for those desiring to review completed research in this area. The bibliography of doctoral research on aging completed by Moore, Tuchin, and Birren (1971) covers the period 1934 thorugh 1969. In analyzing this research Moore and Birren (1971) noted that less than one quarter of one per cent of all doctoral research dealt with any aspect of the agining. Of these investigations far fewer were concerned with reading and the aged. Doctoral dissertations on aging completed during the period 1972 through 1974 are reviewed by Mueller, Moore, and Birren (1975).

Surveys

The studies included in this section are primarily general discussions of problem areas related to research on the aging. Anderson (1960) presents percentage of biological, medical, psychological, and sociological research related to the aged in 15 countries. The author makes no mention of reading and aging specifically as an area of research interest, but states that the acquisition of skills seems to be of little interest to researchers in any country. He notes that biological and medical concerns have received far more attention than have sociological and psychological. In addition, he feels that there is a great need for cross-natural and cross cultural research by means of standardized survey techniques.

The major social-psychological studies of older people according to Shanas (1958) have dealt primarily with personal and social adjustment in later life. The emphasis in these studies has been on the measurement of this behavior. Little has been done on the role of education and its impact on the personal and social development of the aged.

DeCrow (1975) presents the results of a questionnaire sent to 3,500 private and public agencies on the topic of learning opportunities for older adults. Hobbies and recreation, consumer education, health related subjects, and home life were the most popular subjects currently being offered to the older population in this country.

Journals

While journals from a wide range of academic fields published research related to aging much of this effort is on an infrequent basis. To be included in this section the reference had to be one primarily concerned with various aspects of aging.

The *Journal of Gerontology* is a bi-monthly publication containing research grouped into the following areas: biological and medical sciences, psychology and the social sciences, and social gerontology. This is an especially useful reference because of its complete review of current research in gerontology and geriatrics published in each issue.

Aging is published by the Office of Human Development of the Department of Health, Education, and Welfare. This is a general reference journal with particular emphasis on HEW funding and other government related legislation. Current books and journal articles on aging are reviewed in each issue.

Current Literature on Aging is a quarterly journal which published a variety of articles and reports on general aspects of the aging. The fall issue contains a list of annotated bibliographies about adult education for the older citizen.

READING INTERESTS AND HABITS OF THE AGED

A number of the references reviewed for this study dealt with some aspect of the reading interests and habits of the older reader. Typically, these studies were developed around the use of library facilities with particular emphasis on needed changes in these services.

While not concerned exclusively with the older reader Sharon (1973-1974) presents the results of a national survey of the reading habits of 5,067 adults. He notes that the average person spends almost 2 hours each day in some form of reading activity. The majority of this reading is concerned with newspapers, magazines, books, and job-related material.

The rate at which research-based information was introduced from gerontology to librarianship between 1946 and 1969 was investigated by Kanner (1972). He concluded that while other social issues will continue to influence public libraries, the need of an aging population will have increasing emphasis in the library literature and eventual policies of these institutions.

The problem of how can libraries adjust their services to meet the particular reading needs and interests of the older reader was discussed by Barnett (1961), Buswell (1971), Long (1973), and Schmidt (1972). Each suggests a variety of changes, both at the theoretical and practical levels, that libraries must consider if they want the aged to use their facilities.

Hendrickson and Bames (1964) suggest that institutions of higher education should consider developing both formal and informal learning activities, which would keep older persons "mentally alert, constructively occupied, and socially adjusted." Problems associated with the establishment of a curriculum of this type are presented and discussed.

Basic Research

Most of the published research on reading and the aging involves select and small populations. There are some exceptions, though. The studies reported by Ter Veld (1960) and Havighurst (1960) used census data in their analyses of the reading habits, interests, and preferences of aged persons. The *National Survey of Library Services to the Aging* (1971, 1972) described programs specifically designed for the aging as well as what were judged to be exemplary library programs. Ferstl (1974) summarizes the findings and conclusions of this extensive survey suggesting implications for future changes in library services. In discussing the reading habits and interests of the elderly, Schramm (1969) reviewed 13 research studies which used large N's or national samples. In general, the findings of Ter Veld, Havighurst, and Schramm indicated that middle class older persons read more than older persons in the lower classes.

As was stated above, far less than one quarter of one percent of all the doctoral dissertations done in the United States since 1934 concerned, even in part, any aspect of reading and the aging. A search of the ERIC system (90,000 documents scanned with the descriptors Older Adults, Senior Citizens, Library Services, Reading, Reading Level, Reading Habits, Reading Material, Reading Interests) yeilded only four relevant studies (Moshey, 1972; National Survey of Library Services to the Aging, 1971, 1972; Wolfe, 1963).

Hoar (1960) conducted a descriptive survey of 200 aged persons in Oxford, Mississippi, investigating, among other things, the newspaper, magazine, and book reading habits of these subjects. By declining amounts of time, the respondents viewed television, listened to the radio, read newspapers, magazines, and finally books. He reported data on their manner of reading the newspaper and the reading content they preferred. Over 50% had read no books in the previous six months. The more educated the person was the more time he spent reading.

Moshey (1972) surveyed middle and upper-middle class retired professionals to determine the amount and kinds of reading they engaged in. Approximately 90% of the 50 subjects had read newspapers, magazines, and books during the previous six months. Of those surveyed 74% preferred biographies, 68% preferred travel books, and 62% preferred fiction. The average readability levels of the books listed by the participants were levels 7 to 8, but readability did not seem to be a factor in book selection.

In individual, but related ways, Cassata (1967), Soileau (1972), and Wolfe (1963) each investigated aspects of the reading behavior and adjustment of older persons.

Cassata, in her dissertation, which was a descriptive survey of 177 persons aged 62-92, found no outstanding relationships between nine constructed indices of disengagement and eight constructed indices of mass media consumption.

Soileau used a 3% stratefied random sample of the male aged population of a Louisiana parish as the subjects of his dissertation. He administered to these older men the *Leisure Participation Scale* developed by C.R. Pace of the University of Minnesota. One of his conclusions was that leisure involvement in old age strongly correlated with earlier levels of educational, occupational, civic, and professional involvement.

In his dissertation, Wolfe interviewed a random sample of 251 adults who were 65 and older. Using the *Your Activities and Attitudes* scale developed by Burgess, Cavan, and Havighurst, he selected 100 subjects who achieved the top 50 and the lowest 50 adjustment scores. He then administered to them a free recall schedule, which was designed to determine, among other things, whether their adult reading had dealt with occupation, hobbies, world events, or the "liberal arts" or whether it had been devoted to "non-educational" matters like sports or fiction. By means of a 2 x 2 chi square test (well-adjusted and poorly adjusted, educational and non-educational reading), he found that good adjustment was significantly related to educational reading, but was not related to reading of materials assumed to be recreational or "non-educational."

If a researcher were planning to use a semantic differential to ascertain the attitudes of older persons toward some aspect of reading, he would do well to consult a study by Drevenstedt (1975), who found that a group of less well-educated older persons exhibited a greater tendency to mark fewer intermediate scale positions and more neutral midpoint positions than a group of retired professors emeriti and a group of university undergraduates. If a researcher were planning to use interviewers to study older persons, he should consider Freitag's (1972) study of interviewer effects on the responses of 895 aged persons in 29 counties in Georgia. Freitag found that aged persons who had favorable attitudes toward the aging (as determined by their scores on a semantic differential) obtained fewer Not Ascertained responses on a social indicators survey than interviewers who were college faculty or university students.

Conclusion and Implications for Further Research

Research concerning reading and the aging is unchartered territory. The bulk of what little research there is seems to be in the area of reading habits and preferences. Reading's relationship to the adjustment-disengagement of older persons has not been adequately explored. Our findings agree with those of Kingston (1973:55), that little or no research has been carried out on the reading abilities (e.g., word recognition, comprehension) of the aging or on efforts to improve their reading abilities.

The following areas are among the many which merit investigation:

1. the etiology and characteristics of avid and mature aged readers.
2. the flexibility (rate and comprehension) of reading behaviors on the part of aged persons.
3. the construction and validation of instruments designed to assess the reading competency and maturity of aged persons.
4. the effects of bibliotherapy for the aging.
5. the usefulness of readability formula in recommending reading to aged individuals.
6. the relationship between listening comprehension and reading comprehension in aged persons.
7. the relationships between the self-concepts of the aging and their reading behaviors.
8. the critical and creative reading behaviors of the aging.
9. the effectiveness and value of individualized reading programs versus group programs for aged persons.
10. the validity and usefulness of cloze, informal reading inventories, and standardized reading tests for the aging.
11. the relationships between the oral language performance of aged persons and their reading performance.
12. the reading behaviors of important groups of aged persons, e.g., the intellecutally gifted, cultural minorities.

References

Aging. Washington, D.C.: Administration on Aging, Office of Human Development, Department of Health, Education and Welfare.

Anderson, J.E. Research on aging. In Burgess (1960, 25-33).

Barnett, A.N. Beyond librarianship: a critique of rationales of special library service to the aged. *The Library Quarterly,* April 1961, 31, 178-186.

Burgess, E.W. (ed.) *Aging in western societies.* Chicago, University of Chicago Press, 1960.

Buswell, C.H. Reading and the aged. *Wilson Library Journal,* January 1971, **45,** 467-476.

Cassata, M.B. A study of the mass communications behavior and the social disengagement behavior of 177 members of the Age Center of New England. (Doctoral dissertation, Indiana University, 1967) 2765A-3766A. Ann Arbor: University Microfilms. Order No. 68-2275.

Clifton, C.S. A study of the leisure time interests and activities of old age assistance receipients in commercial resthomes in Minnesota. (Doctoral dissertation, University of Minnesota, 1951)

Current Literature on Aging, Washington, D.C.: National Council on Aging.

Darling, F.E. A leisure time analysis of retired public school teachers in Kentucky. (Doctoral dissertation, University of Indiana, 1959)

DeCrow, R. *New learning for older Americans: an overview of the national effort.* Washington, D.C.: Adult Education Association of the United States, 1975.

Drevenstedt, J. Scale-checking styles on the semantic differential among older people. *Journal of Gerontology.* 1975, Vol. 30, No. 2, , 170-173.

Education for aging bibliography. Montclair N.J.: National Multimedia Center for Adult Education, 1975.

Facts about older Americans – 1975. U.S. Department of Health, Education and Welfare. Office of Human Development, Administration on Aging. (OHD) 75-20006, 1975.

Ferstl, K.L. Indiana libraries and services to the elderly. *Focus on Indiana Libraries,* Winter 1974, **28,** 14-18.

Freitag, C.B. Interviewer effects in a social indicators survey of the aged. (Doctoral dissertation, University of Georgia, 1972). *Dissertation Abstracts International,* 5299A, Ann Arbor,: University microfilms, Order No. 73-5694.

Havighurst, R.J. Life beyond family and work. In Burgess (1960, 299-349)

Haworth, E.H. The expressed educational needs of older-age adults in two senior centers. (Doctoral dissertation, Stanford University, 1955)

Hendrickson, A. & Barnes, R.F. *The role of colleges and universities in the education of the aged.* Center for Adult Education, College of Education, The Ohio State University, December, 1964. Cooperative Research Project No. 0E1530.

Hoar, Jr. The reading, listening, and viewing behavior of the aged: an inventory of the mass communications habits and preferences of 200 aged persons in Oxford, Mississippi. (Doctoral dissertation, State University of Iowa, 1960) 4659-4660. Ann Arbor: University microfilms, L.C. Card No. Mic. 60-1556.

Jackson, E.L. Leisure time activities of Negro men, ages 45 through 54. (Doctoral dissertation, Pennsylvania State University, 1955)

Journal of Gerontology. Washington, D.C. Gerontological Society.

Kaplan, J. & Aldridge, G.T. (Eds.) *Social welfare of the aging*. New York: Columbia university press, 1960.

Kannar, E.E. The impact of gerontological concepts on principles of librarianship. (Doctoral dissertation, University of Wisconsin, 1972) *Dissertation Abstracts International*, 32, 5814A. Ann Arbor: University Microfilms, order No. 72-4281.

Kingston, A.J. Areas of concert about adult reading. National Reading Conference 22nd Yearbook, Vol. II. Boone, North Carolina: The Conference, 1973.

Long, F. Research in the field of aging and its relationship to public library services. *Library Trends*, January 1973, 21, 435-439.

MacLean, J.R. An analysis of leisure time activities of selected aged residents of Bartholomew county, Indiana. (Doctoral dissertation, University of Indiana, 1959)

Moore, J.L. & Birren, J.E. Doctoral training in gerontology: an analysis of dissertations on problems of aging in institutions of higher learning in the United States. *Journal of Gerontology*. 1971, Vol. 26, No. 2, 249-257.

Moore, J.L., & Tuchin, M.S. & Birren, J.E., A Bibliography of doctoral dissertations on aging from American institutions of higher learning. 1934-1969. *Journal of Gerontology*. 1971 Vol. 26, No. 3, 391-422.

More Words on Aging Supplement – 1971. U.S. Department of Health, Education, and Welfare. Social and Rehabilitation Service, Administration on Aging. AoA Publication No. 216-S, June, 1971.

Moshey, K.M.S. The retired adult reader: his reading interests and choices and the readability levels of them. May 1972. Ed. 064 674.

Mueller, J.E., Moore, J.L., & Birren, J.E. A bibliography of doctoral dissertations on aging from American institutions of higher learning, 1972-1974. *Journal of Gerontology* Vol. 30, No. 4, 484-489.

National Survey of Library Services to the Aging. Cleveland Public Library, U.S. Department of Health, Education and Welfare, Office of Education, Bureau of Libraries and Educational Technology, December, 1972. ED072835.

Riley, M.W., Riley J.W., & Johnson M.E. (Eds.) *Aging and Society, Volume Two: Aging and the Professions*. New York: Russell Sage, 1969.

Schmidt, L. The aging and the aged. *Library Occurrent*, February 1972, 24, 7-12.

Schramm, W. Aging and mass communication. In Riley *et al.* (1969, 352-375).

Shanas, E. Some sociological research findings about older people pertinent to social work. In *Toward Better Understanding of the Aging* (1958, 49-58)

Sharon, A.T. What do adults read? *Reading Research Quarterly*, Number 2, 1973-1974, 9, 148-169.

Silvey, R.B. Participation in a senior citizens day center. In Kaplan and Aldridge (1960, 88-96).

Soileau, R.R. The implications of early and career socialization for social and leisure pattern among the elderly. (Doctoral dissertation, Louisiana State University and Agricultural and Mechanical College, 1972) 7-47A-7048A. Ann Arbor: University microfilms, Order, No. 73-13, 683.

Ter Veld, J. Pattern of Leisure-time behavior of the aged in the Netherlands. In Kaplan and Aldridge (1960, 139-147).

Toward Better Understanding of the Aging. New York: Council of social work education, 1958.

Words on Aging – A Bibliography. U.S. Department of Health, Education, and Welfare. Social and Rehabilitation Service Administration on Aging. AoA Publication No. 216-A, October, 1970.

Wolfe, L.M. A study of the relationship between lifelong learning and adjustment of older people. (Doctoral dissertation, University of Michigan, 1963) 603. Ann Arbor: University microfilms, Order No. 63-5035.

GRAMMATICAL USAGE AND READING: A REVIEW OF THE LITERATURE

MARY F. FRANKE
University of Kentucky

EARL F. RANKIN
University of Kentucky

In the recent past considerable doubt existed concerning the value of learning "grammar" for the improvement of any aspect of communication, i.e. listening, speaking, reading, or writing. Although countless thousands of English teachers fervently believed in the functional value of grammatical knowledge and skill, scant evidence existed to support their undocumented faith. However, this was in part due to a naive concept of grammar influenced by the prescriptive grammar taught in schools. This naive concept also influenced the operational definitions of grammar in early experimental attempts to demonstrate its utility in facilitating communication skills.

The current influence of linguistics science on both educational practice and research has stimulated more sophisticated grammatical studies. It is the purpose of this paper to review recent literature on the relationship between grammatical usage and reading.

Syntax and Reading Achievement

A number of studies were found to indicate the importance of syntax to reading achievement. Strickland (1962) identified and described patterns of children's oral language. She found that, although children use a large number of patterns, some are used with greater frequency at all grade levels. Ruddell (1965) constructed cloze comprehension passages written with Strickland's high frequency patterns and others with low frequency patterns. His fourth-grade subjects scored significantly higher on the high-frequency tests than on the low. Ruddell was able to conclude that reading comprehension is a function of the similarity of language patterns in textual material and in children's language. On a task involving matching sentences with pictures Tatham (1968) found that second and fourth graders scored significantly higher on a form constructed with frequently used sentence patterns than on an equivalent form constructed with infrequently used sentence patterns.

Granowsky (1971), using second-, third-, and fourth-grade subjects, found that syntax is a significant factor related to readability and that syntactic structures not found in the language of the children were difficult for them to understand in reading. After syntactically analyzing reading errors at first-grade level, Clay (1968) concluded that the "young child's guesses at points of uncertainty in this reading tend to be dominated by his control of the syntax of his lanugage" (p. 437). At the word recognition level dialectal differences in grammar may deny the reader the use of context clues in determining a word not easily recognized by him. Stewart (1966) says:

> He [the reader] can tell generally what the gist of a sentence is if it is close to his own syntax. If it's not close to his syntax, then he has to define what it means and he can't concerntrate on the sound-meaning word correspondences. His problem-solving activities are too diffused. (p. 64)

Weinstein and Rabinovitch (1971) compared the performance of superior fourth-grade readers with the performance of fourth-grade significantly poorer readers on nonsense-word repetition tasks. Noun, verb, adjective, and adverb stems were deleted from four English sentences. Four structured "sentences" were prepared by inserting nonsense syllables into the empty slots, thus retaining structure words, order, and inflections to signify structural meaning. Four unstructured "sentences" were prepared by randomly rearranging the order of the items in each structured "sentence". From the "sentences", four lists were prepared to vary the order, with two structured and two unstructured on each list. Each child heard the list played on tape. His task was to repeat each "sentence" just as he heard it, with three trials permitted for each. After the effects of IQ were partialled out, a significant difference was found between the good readers' performance on structured and on unstructured "sentences", but no significant difference was found in the performance of the poorer readers on structured and unstructured "sentences". These findings suggest that the poorer readers were making little or no use of syntax.

Denner (1970) measured the syntactic competence of samples of first-grade average readers, of first-grade poor readers, of third- through fifth-grade problem readers, and of children in a Head-Start program who were expected to be problem readers. The subjects were taught to associate words with line-drawing symbols, then presented with three of the symbols in a sequence which formed a sentence. The children were to perform the act indicated by the symbol sequence. The first-grade average readers were significantly superior in their performance to each of the two groups of poor readers and the Head-Start subjects. Both the problem readers and the Head-Start children attempted to act out the individual symbols rather than the string of symbols as a unit, indicating that the children were not making use of word relationships to give meaning. All the children, however, were able to communicate orally in sentences.

When Little (1974) compared the oral reading substitution errors of average and disabled third-grade readers, he found average readers' substitutions to be significantly more acceptable both syntactically and semantically than were the disabled readers' substitutions.

Although not a reading study, Epstein's (1961) work has important implications for this review. In order to study the role of syntactical structure on verbal learning, Epstein (1961) constructed a set of "sentences" composed of nonsense syllables together with function words (articles, prepositions, conjunctions) and English inflections appended to simulate standard English structure. A second set of "sentences" was composed of the same nonsense syllables and function words in the same order, but without the inflections. The third set employed the material in the original set, but in random order. A fourth set was constructed of the original nonsense syllables and function words in the same order, but with the inflections shifted to form a competing syntactical pattern not typical of standard English. Sets five and six were composed of meaningful words arranged so that in association they were essentially meaningless. In the fifth set the words were ordered to conform to English structure. In the sixth set the words of the fifth set were presented in random, unstructured order. Epstein's college-student subjects were asked to write each "sentence" after viewing it for seven seconds, with exposures continued until the subject was able to recall the "sentence" perfectly.

Set 1 "setnences" were recalled significantly more easily than set 2 and set 3 "sentences". Contrary to expectations, there was no significant difference in recall of sets 1 and 4. Epstein theorized that the subjects were able to restructure the inflectional patterns. Set 5 was significantly more easily recalled than set 6. In a later study with different subjects and different materials constructed in a similar way, Epstein (1962) was able to replicate his findings. On recall tasks, Rosenberg (1966) also found that the recall of subjects exposed to grammatical sequences was significantly superior to the recall of subjects exposed to ungrammatical sequences. These findings suggest that structure facilitates recall of printed material.

A number of studies were found reporting substantial correlations between syntax and reading achievement. O'Donnell (1963) found a correlation of .46 between grammar test scores on senior high school students and their scores on a test of reading comprehension. Kuntz (1975), in one school, found that seventh graders' scores on a standardized reading test correlated significantly (r=.68) with their total scores on a written sentence-construction test. In a second school the correlation was .80. The sentence-construction test measured the ability to transform two simple sentences into one, employing either (1) prepositional phrases used as noun and adverbial modifiers, (2) appositives, (3) infinitives, or (4) subordinate clauses in noun slots.

Ruddell (1967) tested first graders using a syntax test which required selection of an oral grammatical contrast to match a picture. The grammatical contrasts included singular/plural, present progressive/past tense, present progressive/future tense, simple present/future tense, affirmative/negative, indirect/direct object, subject and object in the active voice, and subject and object in the passive voice. Several months later a written test, similar to the oral one, was administered, which was designed to measure sentence comprehension. Tests of paragraph meaning and vocabulary were also administered. Significant positive correlations between performance on the early test of syntax with sentence comprehension, paragraph comprehension, and vocabulary achievement were found to be .68, .41, and .67, respectively.

Harris (1974) administered a written sentence-construction test to second graders. Parallel items were administered as an oral test of syntax. The correlation of the sums of the written and oral syntax test scores with the scores on a standardized reading test was .70. Torrey (1972) found that second graders' test scores on a standardized reading test correlated significantly with ratings of standard grammatical patterns in sponteneous speech (r=.41) with oral comprehension and production, (r=.50), with use of three grammatical features (inclusion of the copula and articulation of the -s of the third person singular and of the possessive of the noun), (r=.63); and with matching pictures and written phrases, (r=.57).

Inflections and Reading

Four studies were found involving inflections and reading ability. Brittain (1970), using middle-class white subjects, with the effects of IQ partialled out, found a significant correlation of .36 between reading achievement and inflectional performance at first-grade level. At second-grade level, however, she found the correlation to be .70, accounted for, perhaps, by the more meaningful material at second-grade level. While one cannot assume cause and effect from correlations, it is evident that there is a significant relationship between reading achievement and inflectional performance, at least at the second-grade level. Weinstein and Rabinovitch's experimental study discussed in the syntactical section, produced evidence in support of Brittain's findings. Epstein, also discussed previously found that sentences composed of nonsense words with appropriate inflections were recalled significantly more easily than those with no inflections.

Labov (1970) charted the scores of his black 10- to 17-year-old subjects from a 3-item **-ed**-comprehension test to compare those scores with their reading scores, which ranged from 3.9 to 7.0. From examination of the chart he concluded that "successful use of the **-ed** suffix . . . did not correlate with general reading ability." When Labov grouped his subjects according to their scores on the **-ed**-comprehension test and noted that the reading test means of the groups were nearly identical, 4.6 or 4.7, he cited this as further evidence that there is "no correlation between general reading skill and the ability to decipher the **-ed**-suffix as a signal of the past tense." In addition to the shortness of the **-ed**-comprehension test, all scores hinged on the subjects' ability to correctly pronounce the homograph **read** appearing in a sentence with another verb marked by the **-ed** inflection, i.e., When I pass**ed** by, I **read** the posters. Because this part of Labov's study is not based on empirical evidence, it cannot be cited as a valid indication of no relationship between understanding of inflections and general reading ability.

These studies suggest that inflectional performance is related to reading ability. However, more work is needed concerning the specific nature of this relationship.

Nonstandard Speech and Reading

In a study of dialect differences in the speech of first-grade black children, Wiggins (1970) concluded that nonstandard oral grammatical usage was found in greater proportion in the lower reading group than in the upper reading group, although some good readers employed a high incidence of non-standard grammatical patterns in their speech.

Johnson (1970) analyzed taped language samples of black first graders for non-standard grammatical features. One purely phonological feature was also counted **d** substituted for **th.** Scores from a grammar test and a non-sense-word task were also used to compute percentages of nonstandard grammatical usage. Scores from the reading tests were found to have highly significant negative correlations with scores from the grammatical tests for the low-socioeconomic, but not for the high-socioeconomic children.

Deffenbach (1972) found no differnece between the spontaneous speech patterns of high and low black inner-city readers in grades 3 through 6. However, a nonsense-word test revealed significant differences between high and low readers in formulation of noun plurals, possessives of the noun, third person singular forms, and past tense structure, with the high readers using significantly more standard grammatical forms. The findings suggest that the high readers may not consistently use standard English patterns, but are significantly more able to produce such patterns on demand.

Goodwin (1972) developed an oral sentence repetition test composed of sentences likely to elicit nonstandard grammatical features of black non-standard dialect from black and adult basic-education students who were users of a black nonstandard dialect. The dialect test was scored according to the number of standard grammatical features produced. The subjects' dialect test scores were found to positively correlate with their reading scores ($r = .513$).

Franke (1975) found that on a standard English cloze comprehension test, the group of fifth-grade black students who could both produce and comprehend standard English were significantly superior to the group who had difficulty both in producing and comprehending standard English. However, the subjects who had difficulty were not benefitted in reading comprehension by the translation of materials to their own grammatical patterns of speech.

Significant correlations were found between level of nonstandard usage and reading ability. A significant difference was found between high-readers' ability to produce specific standard grammatical forms and that of low readers.

Specific Grammatical Sources of Reading Difficulty

Two studies were found which attempted to pinpoint specific grammatical difficulties contributing to poor reading. Using college students and faculty as subjects in a sentence learning task, Gladney and Krulee (1967) found that grammatical errors in the main verb disrupted reading comprehension significantly more than errors which occurred either in subject or in object position. Errors in the verb produced uncertainty about the ending of the noun phrase and the beginning of the predicate.

Stoodt (1972) found a correlation of .415 between reading comprehension scores of fourth-grade students and a multiple-choice comprehension-of-conjunctions test. With textual readablity held constant, three cloze tests were developed to measure comprehension of conjunctions, one with a high number of conjunctions, another with half as many conjunctions, and the third with no conjunctions. The high-conjunctions cloze test was found to be significantly more difficult than either of the others. The test with half as many conjunctions was significantly more difficult than the one with no conjunctions. The significantly difficult conjunctions were **when, so, but, or, where, while, how, that,** and **if.**

Instruction in Grammatical Patterns

Some students who use nonstandard patterns have shown improvement in reading comprehension when appropriately instructed in syntax. Reed (1968) reported that a group of bi-lingual (Spanish-English) seventh graders, after 30 lessons in sentence and paragraph structure, made significantly greater gains in reading comprehension than did a control group of Spanish-English–speaking seventh graders who were instructed in the conventional English classes. Monolingual students made no such gains.

Stedman's (1971) fourth-grade experimental group was taught to combine kernel sentences into more syntactically difficult forms, while the control group was taught by the more traditional curriculum. On a reading comprehension test the mean of the black experimental group was found to be significantly superior to the mean of the black control group. There was no significant difference in the performance of the two groups of white students.

It is unwise to generalize from the small number of studies reviewed in the last two sections of this paper. More such investigations are needed.

REFERENCES

Brittain, M. M. Inflectional performance and early reading achievement. *Reading Research Quarterly*, 1971, *6*, 34–48.

Clay, M. M. A syntactic analysis of reading errors. *Journal of Verbal Learning and Verbal Behavior*, 1968, *7*, 434–438.

Deffenbach, S. A. Investigation of the oral language patterns of black, inner-city, good, average, and poor readers (Doctoral dissertation, Cornell University, 1972). *Dissertation Abstracts International* 1973, *33*, 5030A. Ann Arbor: University Microfilms, Order No. 73–6, 643.

Denner, B. Representational and syntactic competence of problem readers. *Child Development*, 1970, *41*, 881–887.

Epstein, W. The influence of syntactical structure on learning. *American Journal of Psychology*, 1961, *74*, 80–85.

Epstein, W. A further study of the influence of syntactical structure on learning. *American Journal of Psychology*, 1962, *75*, 121–126.

Franke, M. F. The effect of the grammatical features of their dialect on the reading comprehension of speakers of a nonstandard black dialect of English (Doctoral dissertation, University of Kentucky, 1975).

Gladney, T. A. & Krulee, G. K. The influence of syntactic errors on sentence recognition. *Journal of Verbal Learning and Verbal Behavior*, 1967, *6*, 692–698.

Goodwin, S. H. Development and evaluation of a sentence-repetition task of dialect for black adult basic education students (Doctoral dissertation, North Carolina State University at Raleigh, 1972). *Dissertation Abstracts International*, 1973, *33*, 4043A–4044A.

Granowsky, A. A formula for the analysis of syntactic complexity of primary grade reading materials (Doctoral dissertation, University of Pennsylvania, 1971). *Dissertation Abstracts International*, 1972, *32*, 1746A. Ann Arbor: University Microfilms, Order No. 71–25, 534.

Harris, M. McD. Second grade syntax attainment and reading achievement (Doctoral dissertation, University of Pittsburg, 1974). *Dissertation Abstracts International*, 1975, *35*, 7635A. Ann Arbor: University Microfilms, Order No. 75–13, 189.

Johnson, D. K. An investigation of the oral language and oral reading of black first grade children (Doctoral dissertation, The Ohio State University, 1970). *Dissertation Abstracts International*, 1971, *31*, 94A. Ann Arbor: University Microflims, Order No. 71–18, 028.

Kuntz, M. H. The relationship between written syntactic attainment and reading ability in seventh grade (Doctoral dissertation, University of Pittsburg, 1975). *Dissertation Abstracts International*, 1975, *36*, 2159A. Ann Arbor: University Microfilms, Order No. 75–22, 449.

Labov, W. The reading of the –ed suffix. In H. Levin and J. P. Williams (Eds.), *Basic Studies on Reading*. New York: Basic Books, Inc., 1970, 222–245.

Little, L. J. A study of the relationship between syntactic development and oral reading substitution miscues of average and disabled third grade readers (Doctoral dissertation, University of Kansas, 1974). *Dissertation Abstracts International*, 1975, *35*, 5971A. Ann Arbor: University Microfilms, Order No. 75–6139.

O'Donnell, R. C. A study of the correlation between awareness of structural relationships in English and ability in reading comprehension. *The Journal of Experimental Education*, 1963, *71*, 313–316.

Reed, E. E. Improving comprehension through study of syntax and paragraph structure in seventh-grade English classes. In J. A. Figurel (Ed.), *Forging Ahead in Reading*. Newark, Del.: International Reading Association, 1968, *12* (1), 575–579.

Rosenberg, S. Recall of sentences as a function of syntactic and associative habit. *Journal of Verbal Learning and Verbal Behavior*, 1966, *5*, 392–396.

Ruddell, R. B. The effect of the similarity of oral and written patterns of language structure on reading comprehension. *Elementary English*, 1965, *42*, 403–410.

Ruddell, R. B. Variation in syntactical language development and reading comprehension achievement of selected first grade children. In J. A. Figurel (Ed.), Vistas in reading. *Proceedings of the Eleventh Annual Convention of the International Reading Association*, 1967, *7*, (1), 420–425.

Stedman, N. A. The effect of a curriculum teaching syntactic embedding upon the reading comprehension of fourth grade students (Doctoral dissertation, The Florida State University, 1971). *Dissertation Abstracts International*, 1972, *32*, 4850A-4851A. Ann Arbor: University Microfilms, Order No. 72–10,048.

Stewart, W. A. Nonstandard speech patterns. In A. M. Broening (Ed.), *Baltimore Bulletin of Education*, 1966, *43*, 52–65.

Stoodt, B. L. D. The relationship between understanding grammatical conjunctions and reading comprehension. *Elementary English*, 1972, *49*, 502–504.

Strickland, R. G. The language of elementary school children. Bloomington, Ind.: Indiana University, 1962. (ERIC Document Reproduction Service No. ED 002 970).

Tatham, S. M. Reading comprehension of materials written with select oral language patterns: a study of grades two and four (Doctoral dissertation, University of Wisconsin, 1968). *Dissertation Abstracts International*, 1969, *29*, 4393A–4394A.

Torrey, J. W. The language of black children in the early grades: studies on developing competence in standard English. New London, Conn.: Connecticut College, 1972. (ERIC Document Reproduction Service No. ED 067 690).

Weinstein, R., & Rabinovitch, N. S. Sentence structure and retention in good and poor readers. *Journal of Educational Psychology*, 1971, *62*, 25–30.

Wiggins, A. V. A study of dialect differences in the speech of first grade Negro children in the inner city schools of Cleveland, Ohio (Doctoral dissertation, Indiana University, 1970). *Dissertation Abstracts International*, 1971, *31*, 5682A-5683A. Ann Arbor: University Microfilms, Order No. 71–11, 356.

PREDICTION OF ACADEMIC SUCCESS OF SPECIAL GROUPS OF UNIVERSITY FRESHMEN

JOSEPH A. MUIA
Madison College

EMERY P. BLIESMER
Pennsylvania State University

Many colleges and universities across the country, including the Pennsylvania State University, have initiated "Open Admissions" policies to provide educational opportunities for special groups of college freshmen. Stanley (1971) indicates that students who are admitted under these open admissions policies can be described as those individuals . . . who on the basis of all available information including high school grades, test scores, socio-economic status, race, ethnic origin, and available financial support are likely to have appreciably more academic difficulty than the typical minimally admissable student (p. 207).

A study of the research on college and university admission policies clearly reveals an increasing effort to recruit special groups of students into colleges (Sedlacek, Brooks, & Horowitz, 1972; Sedlacek & Mindus, 1973).

This trend has been accompanied by an increading need to identify the variables that play a significant role in predicting academic success. The identification of these variables for diverse groups is important in determining the effectiveness of special tests in the acceptance and placement of these students in special (compensatory) programs.

The findings of research conducted to determine the predictive validity of Scholastic Aptitude Test (SAT) scores and similar types of test scores, as well as other standardized measures, have been somewhat contradictory. Clark and Plotkin (1963) found that the SAT and similar type tests were of little use in the prediction of the academic success of Black students. However, in her analysis of entrance tests data, Cleary (1966) found that in two out of three integrated colleges the use of SAT scores and combinations of high school grades predicated college grades as well for Blacks as for non-Blacks. Linn (1967), when reexamining data used by Cleary, obtained results substantiating those found by Cleary by using only SAT scores as predictors.

Temp (1971) used SAT scores from 13 integrated institutions in which 300 or more Blacks were enrolled. He found that separate prediction equations must be used to predict grade point averages for Blacks and for Whites, substantiating the findings of Cleary (1968), of Stanley and Porter (1967), and of Kallingal (1971). Thomas and Stanley (1969) found that high school grades do not consistently have the greatest predictive validity in predicting college grades of Black students, particularly of men, whereas they do for Whites.

The foregoing facts, along with the enrollment of special groups into compensatory programs and the increasing attrition rate of these students, establishes the importance of better identifying those variables that influence a student's academic success in a college program and, consequently, his persistence to the attainment of a baccalaureate degree. With these facts in mind, therefore, it was the intent of the investigator to focus primarily in this study of the identification of the preregistration variables that best predict the academic success (grade point average) of a Black group and a White group of specifically admitted students after their enrollment in a ten-week compensatory program and four non-remedial terms following this program.

Subjects of Study and Remedial Program Involved

The subjects used in this study being reported were freshmen students admitted into the 1973 Summer Intensive Remedial Program at the Pennsylvania State University. Financial support for this program was provided under the Higher Educational Equal Opportunity Act 101. The 139 students on whom data were collected came from Pennsylvania high schools in the Wilkes-Barre, Scranton, Pittsburg, Philadelphia, and Harrisburg areas. Of the 139 students, 98 (70.5%) were male and 41 (29.5%) were female, and 95 (68.3%) were White and 44 (31.7%) were Black. The average SAT Verbal and Mathematics scores for the Black group were 347 and 371, respectively; for the White students, the score were 385 and 422, respectively.

These students were involved in a 10 week remedial program which provided instruction in the areas of reading, English, speech, study skills, and mathematics. A grade point average for the summer work was computed for each student. Instruction in the English and communication skills class included handwriting and speaking skills, with emphasis placed on improvement of grammatical skill as well as a student's ability in written composition. The mathematics classes provided the students with instruction related to arithmetic, elementary algebra, college algebra, trigonometry, and calculus, depending upon the needs of each student. Students who performed work that was consistent and commensurate with work done in the regular university mathematics and English programs were awarded credit for the corresponding courses. The reading and study skills classes presented the students with a systematic program designed to increast their overall study skills and to improve the basic reading skills of vocabulary, comprehension, and reading rate.

To assist the students with their social and personal needs, a counseling component was also an integral part of the program. Staff for this component worked closely with the academic area staff and counseled students with personal and social problems which, if left unattended, might adversely affect a student's performance in the classroom.

Variable Data Obtained

Prior to and during each student's enrollment in the special (remedial compensatory) program, various preregistration data were collected and used as independent variables in this study. The following is a list of the independent variable data obtained:

1. Scholastic Aptitude Test (SAT), Verbal and Mathematics scores.
2. Nelson-Denny Reading Test, Vocabulary and Comprehension scores.
3. The Pennsylvania State University English Objective Placement Test raw scores.
4. The Pennsylvania State University English English Essay Placement Test raw scores.
5. Additional data obtained from admissions forms, students' transcripts, and reference sources.
    a. Chronological age.
    b. Sex of subject.
    c. Ethnic background.
    d. Minority/majority status. (This was the relative standing of the student's ethnic group in the high school attended.)
    e. High school enrollment.
    f. Teacher-pupil ratio (This was based on the total number of teachers and pupils in the high school attended.)
    g. High school rank.
    h. Community size. (This was based on the size of the community in which the student resided.)
    i. High school grade point average.

Grade point averages computed for the remedial term and the four following non-remedial terms were used as criterion or dependent variables.

Analyses of Data

Pearson product-moment correlation coefficients were computed among the preregistration variables, using a pairwise deletion for missing cases. Multiple regression coefficients were also computed by regressing the preregistration variables on the grade point averages. A stepwise mode of inclusion in conjunction with the pairwise deletion program for handling missing data was used, with an alpha level of .05. (Nie, Bent, & Hull, 1975.)

Results and Discussion

General Findings

Before describing the results obtained, some general findings will be discussed. Among the general results found were that the attrition rate of the students steadily increased as they moved from the remedial term to successive non-remedial terms when they were enrolled in regularly university courses. At the end of the fifth term of the study, Fall 1974, only 84 (60.0%) of the 139 students has persisted. Males tended to have a greater attrition rate than females. Within ethic groups, Whites were better persisters than Blacks, with Black males having the highest attrition rate for the entire sample (50%). These findings were slightly inconsistent with those of Astin (1972) who found that women and Blacks had the highest drop-out rates.

There are several reasons which may accounted for the high dropout rate among the low achievers involved in this study. Previous research has shown that a greater proportion of Blacks than non-Blacks depend on scholarships and grants for financial aid (Bayer, 1972). Almost all of the students in this program, both Black and White, depended upon the financial aid package for support. Very often these stipends were late in being approved, and this consequently precluded the enrollment of many of these students. On the other hand, the regularly admitted White students more often could depend on their families to provide financial support, which at least guaranteed sufficient funds for enrollment.

Psychological and sociological reasons for the high attrition rate were also evident. Because of the relatively small number of Black students admitted into the Pennsylvania State University, these students had relatively few opportunities to have fulfilling social relationships. In comparison with the lower achieving White students, it appeared that Blacks more often were excluded from becoming part of the university community. Also evident was the fact that low achievers involved in a highly competitive and academic environment often become discouraged and frustrated. This also frequently led to withdrawl from the university environment.

A comparison of the persistence and attrition rates of Black and White groups showed that Whites had a greater persistence rate than did Blacks. At the end of the Fall term, 1974, 24 of 41 (58.5%) Blacks still persisted in the program compared with 60 of the 86 Whites (69.8%).

Specific Findings

An overall analysis of the stepwise multiple regression for the special group of Black and the special group of White term persisters indicated that the variables which predicted academic success differed over each term for each group. (A summary of results of this analysis will be found in Table 1.) The results of the regression analyses necessitated

Table 1

Correlation Coefficients of the Predictor Variables with Grade Point Averages for the Black and White Students Who Persisted During the 1973 Summer Intensive Remedial Program and Any of the Succeeding Four Non-Remedial Terms

	BLACKS					WHITES				
	GPA 1	GPA 2	GPA 3	GPA 4	GPA 5	GPA 1	GPA 2	GPA 3	GPA 4	GPA 5
AGE	–.26	–.25	–.31	–.27	–.09	.17	–.15	–.19	–.09	–.33**
SEX	–.11	–.14	–.08	–.14	–.12	–.30**	–.05	–.18	–.34**	–.24
MIN/MAJ	–.04	.07	–.11	–.20	–.34	.10	–.01	–.07	.07	.17
HS Enrol	–.07	–.06	–.13	–.20	–.06	.16	–.07	–.23**	–.04	–.04
T–P Ratio	.02	–.13	–.01	–.08	–.04	–.02	–.05	.10	.05	.33**
HSR	.24	.28	.15	.25	–.19	.22**	.26**	.22	.15	.29**
CS	.04	.14	–.05	–.06	.09	.01	.07	–.18	–.25**	–.14
SAT–V	.35*	.38*	.17	.40*	.40	.08	.19	.02	.09	.14
SAT–M	.30	.19	.04	.23	.02	.07	.19	.13	.17	.20
ND–V	.46**	.44**	.20	.34	.28	.07	.15	.13	.05	.20
ND–C	.45**	.41*	.21	.44*	.50*	.09	.07	.04	.11	.10
PSU–EO	.41**	.60***	.24	.34	.42*	.11	.24*	.11	.21	.28**
PSU–EE	.13	.33	.08	.18	.14	.00	.02	–.11	–.03	.00
HSGPA	.28	.37*	.37*	.40*	–.05	.24**	.34**	.19	.20	.25

+ MIN/MAJ:	Minority-majority status
HS Enrol:	High School enrollment
T–P Ratio:	Teacher-Pupil ratio (high school)
HSR:	High School rank
CS:	Community size
SAT–V:	SAT Verbal score
SAT–M:	SAT Mathematics score
ND–V:	Nelson-Denny Reading Test, Vocabulary score
ND–C:	Nelson-Denny Reading Test, Comprehension score
PSU–EO:	English Objective Placement Test score
PSU–EE:	English Essay Placement Test score
HSGPA:	High school grade point average

***Significant at .001 level

**Significant at .01 level

*Significant at .05 level

the computation of separate regression coefficients for the Black and White groups. This finding is consistent with that of Temp (1971) and Kallingal (1971). The overall results also showed that within each group the variables that predicted academic success differed substantially over each term. Equations obtained from the stepwise multiple regression were not only different for the two groups, but in general the variables were also better predictors for the grade point averages for the Blacks than for the White group. The fact that the White group members showed less variation in their grade point averages is a possible explanation for the greater predictive validity for the Black students. Stepwise multiple regressions for the Black groups showed that scores obtained with the English Objective Placement Test and the Nelson-Denny Reading Test, administered prior to student enrollment in the special intensive remedial program, were the strongest predictors for this group for the first and fifth terms, with correlations ranging from .41 to .60. However, for the third and fourth terms high school grade point average showed significiant strength in prediction of grades of Black students, with correlations ranging from .37 to .40. These findings, however, clearly suggest that, while academic aptitude and achievement test scores were the strongest predictors, these could not be used to predict academic success for this group consistently across all terms. The data also revealed that variables measuring verbal ability were the strongest predictors for the first and second terms. One reason that may have accounted for this is the fact that the students were enrolled in heavily verbal-oriented classes for the first and second terms. However, as the students progressed into the regular university courses, influence of these variables leveled off. The predictive validity of the Nelson-Denny Comprehension score also increased during the fourth and fifth terms, with correlations ranging from .44 to .50. This most likely occurred because these students were increasingly enrolled in courses that required reading skill.

SAT scores, which are commonly used to predict grade point averages, accounted for only a very small portion of the variance in the stepwise multiple regression equation. This was probably due to the fact that other variables measuring verbal ability entered the analyses earlier and accounted for large parts of the variance. If the Nelson-Denny Comprehension score had been omitted, the SAT Verbal score might be used in place of the Nelson-Denny Vocabulary score, with as much reliability, for the Black students.

An analysis of the stepwise multiple regression tables for the White sample presented results quite different from those presented for the Black group. A comparison of the results for the White group over each term showed that high school rank and high school grade point average were stronger predictors for this group than for the Blacks for the first and second terms. For the remaining three terms, it appeared that various demographic and biographic variables (i.e., school population, community size, minority-majority status, sex, and age) had the most predictive validity for this group.

Overall, then, the results of this study suggest that, of the variables studied, none could be used as a common predictor of academic success of Blacks and Whites as a combined group. The findings also revealed that the variables that showed predictive strength for each group were not consistent across all terms. This finding suggests that different combinations of variables must be used to predict the grade point average of students who drop out and re-enroll during another term. The findings also showed that the academic aptitude and achievement test scores were effective preditors for Black grade point average for the first term of university enrollment while for the White group high school enrollment and high school grade point average were stronger predictors.

References

Astin, A.Q. College Dropouts: *A National Profile.* Washington: American Council on Education, 1972.

Bayer, A. *The Black College Freshman: Characteristics and Recent Trends.* Washington: American Council on Education, 1972.

Clark, K.B. and Plotkin, L. *The Negro Student at Integrated Colleges.* New York: National Scholarship Service Fund for Negro Students, 1963.

Cleary, T.A. Test bias: Validity of the Scholastic aptitude test for Negro and White students in integrated colleges. Educational Testing Service Research Bulletin RB-66-31. Princeton, N.J.: Educational Testing Service, June, 1966.

Cleary, T.A. Test Bias: Prediction of Grades of Negro and White Students at Integrated Colleges. *Journal of Educational Measurement,* 1968, 5, 115-224.

Kallingal, A. The prediction of grades for Black and White students at Michigan State University. *Journal of Educational Measurement.* 1971, 8, 263-264.

Linn, R.L. Reanalysis of the data of Miss Cleary's predictive bias study. Princeton, N.J. Educational Testing Service, 1967.

Nie, N., Bent, D.H., and Hull, C.H. Statistical Package for the Social Sciences, New York: McGraw-Hill Book Co., 1975.

Sedlacek, W.E., Brooks, G.C., and Horowitz J. Black admissions to large universities: Are things changing. *Journal of College Student Personnel,* 1972, 13, 305-310.

Sedlacek, W.E., Brooks, G.C., and Mindus, L.A. Black and other minority admissions to large universities: Three year national trends. *Journal of College Student Personnel,* 1973, 14, 16–21.

Stanley, J.C. Predicting College Success of the Educationally Disadvantaged. *Science*, 1971, **171**, 640-647.

Stanley, J.C. and Porter, A.C. Correlation of scholastic aptitude test score with college grades for Negroes versus Whites. *Journal of Educational Measurement*, 1967, **4**, 199-217.

Temp, G. Validity of the SAT for Blacks and Whites in thirteen integrated institutions. *Journal of Educational Measurement*, 1971, **8**, 245-251.

Thomas, C.L. and Stanley, J.C. Effectiveness of high school grades for predicting college grades of Black students: A review and discussion. *Journal of Educational Measurement*, 1969, **6**, 203-215.

PREDICTION OF COLLEGE SUCCESS: A TEST OF STANDARD MEASURES FOR TRADITIONAL AND OPEN ADMISSIONS STUDENTS

JOSEPHINE S. GOLDSMITH
Rutgers University

MARK J. NICOLICH
Rider College

Since the inception of open admissions programs in the 60s, writers in the field have questioned the suitability of standard predictive criteria of academic success for non-traditional students. One of the central issues explored has been that of test bias. Although the chief expressed concern relative to test bias appears to be the possibility of under-prediction of minority students (Clark, 1971; Manning, 1971), there are obviously other possible manifestations which need to be explored. Cleary (1968) has suggested two methods of investigating bias: the testing of item-group interaction and the appearance of consistent non-zero errors of prediction for the subgroup.

The results of examinations of test bias are diverse and complex enough to make attempts to draw generalizations hazardous, especially as these might be used to guide admissions policies. However, some farily reliable findings appear to emerge. It has been found, for example, that SAT scores predict roughly equal but unprepossessing amounts of variance for both white and black student groups (Cleary, 1968; Temp. 1971) and that this phenomenon persists in both predominantly black and integrated institutions. The use of regression equations developed upon white or total populations predicts black students to achieve about as well or better than they do (Stanley & Porter, 1967; Stanley, 1971; Temp, 1971). A recent summation of this literature (Cleary, Humphreys, Kendrick, & Wesman, 1975) concludes that standard predictors are useful for both student groups.

These findings, fairly consistent as they are, do not put the question of test bias to rest. Both Cleary (1968) and Temp (1971) investigated the similarity of regression planes of SAT upon GPA for both majority and black students. Cleary, testing equality of slopes and intercepts, found no difference in slopes, but a difference in intercepts in one of the three colleges tested. Temp, using the Gulliksen-Wilks procedure on data from thirteen integrated institutions, found significant differences in 43 of 52 tested regression planes. He concluded that "a common prediction system is not possible and that a separate prediction system would have to be developed for each subgroup" (Temp, p. 249). On the basis of the difficulty of generalizing from his widely disparate findings, Temp makes a strong point for colleges to do their own validity investigations.

While writers have tested the efficiency of traditional predictors for black versus white groups, fewer have extended studies to other minority groups. In 1973 Oppenheim reported a comprehensive validity study for freshmen in five colleges of a large, partially state supported, Eastern university. Data were examined for the "disadvantaged" students in the Educational Opportunity Program as well as for students admitted under standard criteria. The Educational Opportunity Program (EOP) students represented a variety of ethnic groups, with nine ethnicity codes used, including white, black, Hispanic, Oriental, and others. The proportional contributions of SAT–V, SAT–M and high school percentile rank were assessed with the dependent variable of freshman grade point average.

The present study attempted to assess the current status of predictive validity within one selective college of that university. The findings from Oppenheim and the present study report on a wider spectrum of the disadvantaged than is often included in validity studies of this type. While adding some new features, the present study is largely an updating of one sement of Oppenheim's work.

Method

Sample

The EOP sample consisted of the total number of students in the program at Rugers College for whom complete data sets are available. Five students from the EOP at the School of Engineering were also included, since admission criteria and grading standards were closely similar. The total N for the EOP sample was 180. These students were sophomores, juniors and seniors. A random sample of 70 regularly admitted students was drawn from each of these classes for a sample of 210 standard admits (SA). One student was dropped due to improper values which yielded a final sample of 209.

The Educational Opportunity Program, a state supported project, has been in existence at Rutgers College since the mid-60s. Criteria for admission include both educational and economic disadvantage. Because of concern over high attrition rates, the open standards were tightened in Fall 1974 when standing in the fourth decile or above became an admission requirement. Since the present sample spans admissions from '72 through '74 there would be an expected increment in admissions criteria for the last class.

Data Analysis

A standard multiple regression technique was used to determine the relationship between four independent variables upon the dependent variable, grade point average (GPA). The independent variables chosen were SAT Verbal, SAT Math, high school decile rank, and sex. Means, variance, correlations and scatter plots of the variables were also examined.

Results

An examination of the means indicated that the SA group was higher than the EOP, both practically and significantly. The SA students had higher scores in SAT Verbal, SAT Math, high school decile and college cumulative average (GPA). (In the Rutgers grading system A=1.0; D=4.0.) It was also noted that the variances of the EOP group were significantly (less than .01) greater than the SA group, a relation which held for all four measures. This indicated less uniformity in the EOP group.

The correlations with the sex variable (Female = 2; Male = 1) produced different patterns for EOP and SA. The EOP group showed a significant correlation only between sex and high school decile. Since the correlation was negative it indicated that females has a higher decile standing than males. For the SA group there were significant correlations between sex and SAT Math, sex and high school decile, and sex and GPA. There was no significant relationship between sex and SAT Verbal. For the SA students, females received better grades in both high school and college while males did better on SAT Math.

The other correlations showed a similar pattern for EOP and SA students. One notable exception was the relationship between SAT Math and SAT Verbal. Here, the EOP students showed a stronger correlation (.707) than did the SA (.498). The EOP group scores ran almost the full range for both SAT's. In contrast the SA students' scores were more compact at the high end, hence the SA scores were truncated at the low end.

Regression

To assess the relationship among the variables, three separate regression analyses were performed. An equation was fitted with grade point average as the dependent variable while sex, SAT Verbal, SAT Math and high school decile served as independent variables. The fitting was done for the total group (SA and EOP, N=389), for the EOP group (N=180) and the SA group (N=209).

On the basis of the R^2 values (or corresponding F values for significance of regression) it appeared that the equation based on the total group (EOP and SA) yielded the best prediction of college grade point average. But, investigation of the errors of prediction, or error mean square (EMS) indicated the existence of two distinct groups. The SA group was higher on all measures; the EOP group was lower on all measures.

When the SA and EOP groups were considered separately, each was more homogeneous, resulting in lower R^2 values for the separate regressions. Consideration of the EMS, an indicator of the size of the average error of prediction, showed a different pattern. For the SA group the EMS was about 65% of that of the group EMS, while for the EOP EMS was 132% of the total group EMS. Clearly there were two different phenomena occurring. The EMS's of the two subgroups were significantly different at less than the .001 level.

In view of the evidence the two groups must be analyzed separately. The EOP analysis indicated that the only practically important predictor was the SAT Verbal score, but due to its high correlation (.707) with SAT Math, the Math could substitute for the Verbal. Neither high school decile nor sex was important as a predictor. For the EOP group, then, the most useful predictor was either one of the SAT scores. It must be stressed, however, that the SAT score was the best of a group of exceedingly weak predictors, since only 9% of the total group variability was explained by the independent variables.

For the SA group the equation was more complex with the four variables, sex, high school decile, SAT Verbal and SAT Math being important. Each term yielded a different piece of information; taken together they accounted for 28% of the total variability, which was more than three times the explanatory power of the variables for the EOP group.

These analyses were duplicated adding the interactions (cross products) of SAT Math, SAT Verbal and high school decile as independent variables. Since these interactions did not account for significant reduction in the error sum of squares, these terms were not considered important in this analysis.

Conclusions

The relationship between standard admission criteria and grade point average was different for the EOP and the standard admit sample. While it might be statistically valid to use the same regression equation for EOP and standard admit students, the use of one equation covers the evidence of two different relationships. The relations did not differ merely in regression coefficient values, but were intrinsically of a different form. The results delineate a disparate relationship between GPA and traditional admission criteria for the EOP and standard admit students in the samples. For the SA group, GPA could be predicted adequately ($R^2=.28$) based on evidence from four sources each of which provided relevant non-redundant information. For the EOP group, GPA was poorly predicted ($R^2=.09$) based on only one source, the other three being not related or redundant. There must be other, yet unknown, independent variables that relate to GPA for the EOP group. If bias is defined as a different relationship between predictor and criterion variables, the findings clearly indicate bias in SAT Verbal, SAT Math and high school decile standing for the groups tested.

These findings run counter to studies which suggest fair utility of predictors for both traditional and minority groups. It may be that social class differences exist between the EOP students, defined as disadvantaged, and the students in black colleges studied by Cleary and others. There were differences in underlying forms of the regression equations for the two groups. Different patterns appear to be operating for EOP and SA.

More careful analyses, it appears, yield a well founded discomfort with heavy reliance on traditional measures, especially to predict minority success. The College Entrance Examination Board has never recommended verification of its instruments. Angoff has clearly stated that SAT's are most effectively used as additional information to guide admissions decisions (Temp, 1971). While there has been active debate on possible biases within the tests, perhaps attention needs to be redirected toward popular and typical misuse.

Since CEEB suggests the tests be used in a supplementary fashion, the immediate question becomes, "supplementary to what?" Possibly the search for objectivity and efficiency in admission decisions has become one more social circularity, selecting poorly for all, but less feebly for the children of the middle class.

The call for new measures, possibly criterion referenced, or mastery oriented (Cleary, et al., 1975) or examinations of process (Goldman, 1973) has gone out in the field. Obviously the relationships found here need to be replicated. It would appear urgent for colleges to keep a continual check on the validity patterns for various subgroups within the institution. The need seems clear to develop finer nets to catch student potential.

References

Cleary, T.A. Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, 1968, *5*, 115-124.

Cleary, T.A., Humphreys, L.G., Kendrick, S.A., & Wesman, A. Educational uses of tests with disadvantaged students. *American Psychologist*, 1975, *30*, 15-41.

Clark, K.B. Response to Julian C. Stanley's paper. *Barriers to higher education.* College Entrance Examination Board, 1971, 78-80.

Goldman, R.D. Hidden opportunities in the prediction of college grades for different subgroups. *Journal of Educational Measurement, 10*, 205-210.

Manning, W.H. Personal and institutional assessment: Alternatives to tests of scholastic aptitude and achievement in the admissions process. In *Barriers to higher education*. College Entrance Examination Board, 1971, 81-99

Oppenheim, D. Summary of Rutgers University validity studies class entering September 1973. Rutgers University, mimeo.

Stanley, J.C. Predicting college success of educationally disadvantaged students. *Barriers to higher education*. College Entrance Examination Board, 1971, 58-77.

Stanley, J., & Porter, A. Correlation of scholastic aptitude score with college grades for negroes versus whites. *Journal of Educational Measurement*, 1967, *4*, 199-218.

Temp, G. Validity of the SAT for blacks and whites in thriteen integrated institutions. *Journal of Educational Measurement*, 1971, *8*, 245-251.

INTEGRATING READING AND STUDY SKILLS INTO COLLEGE BIOLOGY

BARBARA TOMLINSON
University of California at Riverside

MICHAEL D. TOMLINSON
Riverside, California

ABSTRACT

Many students who encounter the rigorous demands of introductory courses in biology find themselves struggling and disappointed in their first quarter of college work. One possible solution to this problem is the teaching of reading and studying techniques in close conjunction with content area work. An analysis of skills needed to read and understand the material for introductory biology was done by a college reading specialist and a student consultant in the biology department. An adjunct reading and study skills course was then designed to teach these skills in conjunction with their actual use on assigned biology material. The study course focused on such skills as surveying and constructing an overview map, analyzing text material for comprehension, developing vocabulary through context and systematic study, analyzing Greek and Latin roots, understanding diagrams and graphs, and developing skills of self-questioning at the literal, interpretive, and integrative levels. The course attempted not only to introduce these skills, but also to establish habits of active study.

An assessment was made to determine whether the adjunct reading and study course has succeeded in influencing the mid-term examination scores of enrolled students. Students were assigned to one of six adjunct sections according to their class schedules. Three sections were randomly chosen and offered as adjunct study classes. The students originally assigned to the remaining sections served as the control group. Scores on the verbal and mathematical portions of the Scholastic Aptitude Test were found to have been equal for both groups, with the non-significant twelve-to-fourteen-point difference favoring the control group in both cases. Scores of the two groups on the course mid-term were subjected to analysis of variance to determine the effectiveness of the adjunct class. Students enrolled in the adjunct class received significantly higher scores than the control group $F(1,102) = 11.82$, $p < .005$. The mean of the control group, 36.69 (of 66 points), represented a "solid C" grade. The mean of the adjunct

group, 43.95, represented a "borderline B". Evidence, therefore, indicates that courses which attempt to integrate the teaching of reading and study strategies with content area material may be effective in increasing freshman achievement. It is suggested that success may occur when students are able to practice extensively and internalize effective study strategies in the adjunct class setting.

TRANSFERRING READING AND STUDY SKILLS TO THE CONTENT AREA

TERYLE CARPENTER
Clemson University

YVONNE JONES
Clemson University

In recent years there have been several studies reporting the effect of a reading study-skills program upon academic performance. Mynatt (1973) examined a developmental education program in a community college and its effects upon self-concept, grade point average (GPA), and attrition rate. He found that the program did not have a statistically significant effect on self-concept. There were significant GPA differences for persons enrolled in the program as opposed to those advised to enter but who rejected enrollment. Furthermore, the general studies program had greater holding power for its students than did any of the other forms of curricula under investigation.

A study conducted by Sieberman (1973) investigated the time needed for a student to show specific gains and a definite carry over of skills into the content areas. There was no significant difference in the overall grade averages of those in the reading classes and those not in the reading classes.

Tillman (1972-73) reviewed thirty-one four-year college reading programs which emphasized reading improvement and used grade point average as a criterion. These studies were reported in the literautre between 1945 and 1971. Although twenty-three of the thirty-one studies found statistically significant gains in grade point average for the participants, grade gains tended to be small. It was recommended that grades be viewed as only one factor in evaluating reading programs.

A group of University of Connecticut failing freshmen who received one hour each of individual counseling, group guidance, and study skills weekly for ten weeks were found by Kaye (1972) to have received significantly higher grade point averages the second semester than did a control group. His study also found a significantly higher number of them were still in school the following semester.

University of Manitoba students were found by Ritter (1971) to have made significant grade point average gains four to seven months following their participation in a study skills program of five one-hour sessions. These students rated underlining and "developing study sheets" lessons as most useful.

A study at Clemson University by Carpenter and Sawyer (1971) found that a reading improvement program was beneficial to students with limited verbal skills. In addition, the mean grade point ratios of this group were consistently higher and the attrition rate significantly less.

The benefits in grade point ratio and lower dropout level were affirmed by Robert Sawyer (1969) in a study carried out at the University of Missouri at Rolla. On the other hand, Colvin (1968) found no significant differences on grade point ratio or reading achievement between the experimental and control group in his study. He did find, however, that both groups showed a significant growth in reading achievement.

A study by Earl Thomas (1968), with the aid of his English teachers, noted that "the ability to read more effectively" was the major objective of remedial work in English. Thomas pointed out that reading skill was generally accepted to be a factor in increasing academic achievement.

Ikenberry, Kennedy, and Field (1966) found that significant benefits which lasted at least through their freshman year were derived by students in the reading skills program. Two of these benefits were increased grade point ratio and a lower withdrawal rate.

The purpose of the present study was to evaluate the effectiveness of a developmental college reading program in terms of reading gains, grade point ratios, English grades, and history grades during the 1974-75 academic year.

It was hypothesized that no significant differences would exist (.05 level) between:

1. the mean Sat verbal scores, the mean grade point ratios, the mean English grades and the mean history grades of the experimental and control groups for the 1974-75 academic year.
2. the mean pre and posttest reading scores of the experimental group.

Futher, it was hypothesized that there would be no significant correlations (.05 level) between the SAT verbal scores and all variables for either group.

Procedure

Subjects

Subjects included in the experimental group were 50 freshman students enrolled in the developmental reading course in the fall of 1974 at a small Southeastern university. The subjects in the control group were 50 freshmen who requested the course, but were not enrolled due to class size limits.

Materials and Methods

Approximately six weeks were spent in class on individual acquisition of the subskills of comprehension, vocabulary, and rate. Classes met three times a week for fifty minutes during which each subskill was allotted fifteen minutes. The remaining eight weeks were spent on the application of these skills to textbooks, paperback books, and journals. The subskills were presented by a combination of lectures, handouts, and reading improvement textbooks.

Results

An analysis of variance was made between the experimental and control groups to determine whether there were significant differences (.05 level) between groups on SAT verbal scores, grade point ratios, English grades and history grades for the 1974-75 academic year. Also, an analysis of variance was made between the pre and post reading test scores of the experimental group to ascertain if there were significant gains (.05 level). Finally, a test for correlation between SAT verbal and all variables was made to determine if there were any significant correlations (.05 level).

Results indicated significant differences between the experimental and control group on SAT verbal scores F (1,99) = 47.83, $p < .01$; between the experimental and control groups on English first semester F (1,95) = 5.46, $p < .05$ and between the experimental and control groups on grade point ratio first semester F (1,92) = 12.93, $p < .01$. No significant difference obtained between the experimental and control groups on second semester grade point ratio. SAT verbal scores did not appear to affect first semester grade point ratio significantly, but the group placement did. Significant differences were found between groups on first semester history grades, F (1,34) = 4.97, $p < .05$, but no significant difference obtained on second semester history grades between groups.

Correlation coefficients of SAT verbal scores and all variables for the experimental group were found to be significant. The correlation coefficient of SAT verbal and English first semester and English second semester were significant for the control group but not for the other variables.

Discussion

These data indicate that the transfer of reading skills, gained in a special program, to actual reading experiences in the content areas is likely to occur when direct application of the skills to the content areas is made. Further, treatment was effective in increasing the reading ability of the experimental group.

Conclusion

Additional research is needed to determine if the enrollment of a student in a sequential reading course will continue his significantly better performance during the second semester as well as maintain such performance in subsequent semesters.

References

Carpenter, T. & Sawyer, C., *The effect of a college reading program upon the reading gains, grade point ratio and attrition rate for high risk freshmen.* 1971, ED053901.

Colvin, C.R., "A Reading program that failed – or did it?", *Journal of Reading*, 1968, *12*, 142-46.

Ikenberry, S.O., Kennedy, E.C., & Field, J.V., *Effects of reading, study skills improvement, and reduced credit load on achievement and persistence of failure prone college freshmen: a pilot study*, 1966, ED022654.

Kaye, R.A., The effectiveness of a guidance-counseling study program on the academic achievement of failing college freshmen. (Doctoral dissertation, University of Connecticut, 1971). *Dissertation Abstracts International*, 1 972, *32*, 180A. University Microfilms, No. 71-78.

Mynatt, H.L., The effects of a developmental education program in a community college upon self-concept, grade point average, and attrition. (Doctoral dissertation, East Texas State University, 1972). *Dissertation Abstracts International*, 1973, *33*, 4049A, University Microfilms, No. 73-4442.

Ritter, J.H., University study skills program, *Journal of Reading*, 1971, *14*, 377-380.

Sawyer, R.N., *The effect of specialized developmental reading and study skills instruction and counseling on a sample of students with above average quantitative and below average verbal skills*, 1969, ED031385.

Siebenman, J. et al. *The contribution of reading instruction to success in the content area in the secondary schools*, 1973, ED083546.

Thomas, E. Tentative objectives for freshmen remedial English, *College Composition and Communication*, 1958, *5*.

Tillman, C.E., Four-year college reading improvement programs and grades: an annotated review, 1945-1971, *Journal of Reading Behavior*, Spring 1972-73, *5*, 100-109.

MEASUREMENT OF COLLEGE STUDENTS' SELF CONCEPT: A PILOT STUDY

ANN MARIE BERNAZZA HAASE
University of Massachusetts

MERVIN D. LYNCH
Northeastern University

SHIRLEY DESHIELDS
University of Massachusetts

For the high risk college student, the question of his ability to perform in college is critical. In order to insure success in college, the high risk student usually is required to enter a reading improvement program. It is assumed that exit from such a remedial reading program is evidence not only of academic success but of a heightened self concept (Mangieri & Olsen, 1975). Testing academic success depends on standardized achievement tests used in the reading centers while testing self concept or self worth appears more difficult. Measures of self worth are either cumbersome involving the use of the Q sort or have little to do with the reading situation the student finds himself in. The need was to find a measure of self worth that would be directly related and sensitive to a college student's feeling about himself in relation to reading. A review of the literature revealed studies by Roth (1959) who used the Q sort approach and Estes (1974, 1971) whose attitude scale was reliable for students in grades 3-12 and measured how they felt about personal reading endeavors. Since neither scale adapted itself readily to the college student population, the authors decided to use a semantic differential technique specifically adapted to college students in a reading program. The use of the semantic differential for self concept measurement has been discussed by Wylie (1961).

The purpose of this study was to pilot test the use of this semantic differential instrument with high risk college students in a summer reading/study skills program.

Procedure

Fifty high risk entering college freshman students, with a mean age of 19, enrolled in a private college summer communication skills program and participated initially in the study. Although all students completed the program, only 7 men and 10 women completed both the pre and posttest semantic differential self concept inventory. The subject attrition rate was high due to the voluntary nature of completing the inventory. All students were receiving three credits for the course.

Program

Students were academically pre and posttested with the Iowa Silent Reading Tests. From an analysis of pre-test scores, individual prescriptions were formulated for each student. Emphasis was placed on reading in the content areas skills and on writing skills. Materials used were *Probe*, *Quest: Academic Skills Program*, and the Science Research Associates reading materials.

Semantic Differential Self Concept Inventory (SDSCI)

Ratings were obtained with 14 bi-polar adjective pairs on each of eight concepts: 1) How I feel about: Myself, 2) Myself as I would like to be, 3) Myself when I am reading, 4) Myself in college next semester, 5) Myself when I am writing, 6) Myself when I am studying, 7) Myself when I am in this program, and 8) How others feel about me. Each bi-polar adjectives was evaluated on a scale from 1 to 7. The larger the score, the more positive the evaluation.

The sum score for three adjective pairs comprising the evaluative dimension of meaning for the concept "How I feel about myself," provided a measure of attitudes towards the self or self worth. The sum score of these three items on all other concepts provided a measure of attitudes toward each concept. The D statistic was computed between the concept "How I feel about myself" and each of the other seven concepts. These D statistics provided seven separate dependent measures. These were: personal security with reading (1 vs 3), security with writing (1 vs 5), personal security with studying (1 vs 6), personal security with the remedial program (1 vs 7), personal security with college (1 vs 4), idealized self concepts (1 vs 2), emphatic self concept (1 vs 8). A larger D statistic indicates negative direction. In addition, a measure of self enhancement was obtained by comparing pre and posttest scores on all seven D statistic measures with a Wilcoxon Paired Replicate Test.

Separate analyses of variance were run on each of the seven D statistic measures and each of the eight evaluative sum scores. In each of these analysis of variances, separate comparisons were made between levels of sex and times of testing, pre and post.

Results

Semantic Differential Self Concept Inventory

Three of the analysis of variance on the D statistic measures showed significant results. The difference between how the students felt about themselves and their ideal self increased negatively from pre to posttest scores. Both males and females were further away from realizing their ideal selves at the end of the program than they were at the beginning ($\bar{X}_{pre} = 5.06$, $\bar{X}_{post} = 6.73$, $F = 4.458$, $df = 1, 15$, $p<.05$). No significant difference was found between pre and posttest scores on the difference between "Myself" and "How I feel about myself when I am reading." However, sex differences appeared ($F = 9.298$, $df = 1, 15$, $p<.05$). Men ($\bar{X} = 7.02$) felt less negative about themselves than women ($\bar{X} = 9.25$) who felt extremely negative about themselves when they were reading. Even though no significant differences between pre and posttests were found, the means for both increased, suggesting greater negative affect in the reading situation. An inspection of the means revealed men increased in negative feeling from the pre ($\bar{X} = 6.06$) to the post test ($\bar{X} = 7.99$) while women slightly decreased their negative affect from pre ($\bar{X} = 9.64$) to post test ($\bar{X} = 8.85$). A similar result obtained when the difference between "How I feel about myself" and "How I feel about myself when I am studying" was compared over men and women. Women ($\bar{X} = 9.07$) felt more negative about themselves than men ($\bar{X} = 7.72$), ($F = 5.946$, $df = 1, 15$, $p<.05$). Analysis of the means from pre to posttest for each sex showed men tended toward more negative feelings on both the pre and posttest, while women had more positive feelings at the end of the program, i.e. the women's mean scores were more similar to the men's at the end of the program than at the beginning.

Eight other analysis of variance were performed on the sum scores of the three evaluative items for each of the eight concepts. All analysis of variance were not significant except the analysis of variance for contact two, "Myself as I would like to be" which showed a sex difference ($F = 4.557$, $df = 1, 15$, $p<.05$). Women ($\bar{X} = 4.13$) reported their attitudes towards their ideal selves as more positive than men ($\bar{X} = 4.88$).

In order to assess self enhancement between pre and posttest results, a Wilcoxon Paired Replicate test was performed over the eight D statistic measures. Results showed the direction was significant in the negative direction (the D statistic become larger) for both men ($F = 3$, $df = 8$, $p<.05$) and women ($F = 3$, $df = 8$, $p<.05$). The same analysis was applied to the direction of the eight analysis of variances for the evaluative items; no significant difference in directionality was found for either sex.

The question of why the negative direction in self enhancement arose was asked. Several students who participated in the program were asked to assess their summer communications skills program. The majority commented that they felt it emphasized their weaknesses rather than their strengths, the meaterials were juvenile, boring, and beneath them, and the method of studying and writing taught was long, tedious and time-consuming in that it took too long to apply it to a real studying situation.

Conclusions

The results of this pilot study indicate the semantic differential self concept inventory may be used to assess the affect of students in a reading/study skills program. Presented as a group test, it was sensitive to the slight shift of feelings of each student and to the differences between the sexes. The significant negative self worth feelings supported the evidence of Brown and Newman (1967) who studied adult illiterates and found they have a poorer self concept than their better educated contempories and Roth (1959) who found non-achieving students in reading improvement classes are most resistant to change and perceive the demands placed upon them as threatening to their self esteem. In the face of this threat, the high risk student distorts the experience by criticizing the materials and instructor rather than integrating the experience and changing as a result of it and for his/her benefit.

It is interesting to note also the differences in self worth between the sexes. Women tended to feel more negative about themselves until they realized there were men who also felt unhappy with themselves and who also were resistant to change. The self worth scores of women increased as a result of such an exposure in both the reading and studying situation.

References

Brown, D. A. and Newman, A. Attitudes of adult illiterates toward reading materials and educational programs. In George B. Schick and Merrill M. May (Ed.) *Multidisciplinary Aspects of College-Adult Reading: 17th Yearbook of the National Reading Conference*, Wisconsin: National Reading Conference, Inc., 1968, 167-175.

Estes, T. H. A scale to measure attitudes toward reading. *Journal of Reading*, 1971, 15, 135-138.

Estes, T. H. and Johnstone, J. P. Assessing attitudes toward reading: a validation study. In Phil L. Nacke (Ed.) *Interaction: Research and Practice for College Adult Reading: 23rd Yearbook of the National Reading Conference*. Clemson, South Carolina: National Reading Conference, Inc., 1974, 219-223.

Mangieri, J. N. and Olsen, H. D. Self-concept of academic ability and reading proficiency. In George McNinch and Wallace Miller (Eds.) Reading: Convention *and Inquiry: 24th Yearbook of the National Reading Conference.* Clemson, South Carolina: National Reading Conference, Inc., 1975, 114. (Abstract)

Roth, R. M. The role of self concept in achievement. *Journal of Experimental Education,* 1959, 27, 265-281.

Wylie, R. C. *The self concept.* Lincoln: University of Nebraska Press, 1961.

THE EFFECTS OF TESTWISENESS TRAINING ON READING SCORES AND ON TEST ANXIETY

GLORIA BATKIN KAHN
Brooklyn College of C.U.N.Y.

ABSTRACT

Testwiseness training, which teaches ways to maximize achievement on objective tests, has resulted in significant gains on post-test scores of Testwiseness. However, the transfer of Testwiseness training to subject matter tests has not been studied.

The study sought to investigate the following questions: (a) the effects of Testwiseness training upon standardized reading scores; (b) the effects of Testwiseness training upon test anxiety; (c) the interrelationships between reading speed, reading level and test anxiety.

Sixty-seven Brooklyn College juniors were randomly assigned to the experimental and control groups. The experimental group received four hours of lecture and discussion about testwiseness. The control group received four hours of teaching that focussed on learning theories irrelevant to test taking.

Each group was pre-tested using the Davis Reading Test Form 1A and the Mandler-Sarason Test Anxiety Scale. Three weeks after the training, the groups were post-tested with Form 1B of the Davis and the same form of the Mandler-Sarason. The subjects used pseudonyms to preserve anonymity.

Results were determined by analysis of variance, analysis of covariance, and Pearson correlation coefficients. These results suggested that Testwiseness training had no significant effect upon the ability to take the Davis reading test, but that Testwiseness training did have a significant effect in decreasing test anxiety. There were no significant interrelationships between test anxiety and reading speed, nor between test anxiety and reading level.

Conclusions were that Testwiseness training does not necessarily transfer to a reading test as psychometrically correct as the Davis. It was suggested that Testwiseness training not be considered useful until proof of retention and transfer of training are accomplished. The effects of anxiety upon test taking were unclear. Extremes of test anxiety may not effect reading speed and comprehension. A relationship between reduction of test anxiety and Testwiseness training was suggested.

COLLEGE STUDENTS' PERCEPTIONS OF CHEATING AND NON–CHEATING BEHAVIORS DURING TEST–TAKING

KEN L. DULIN
University of Wisconsin-Madison

M. JANE GREENEWALD
University of Wisconsin-LaCrosse

ABSTRACT

Cheating on school-examinations is apparently on the rise in America's colleges and universities; popular magazines and newspapers regularly report the demise of "honor" codes and the establishment or reinstatement of stricter and more carefully monitored testing situations. This study investigated (1) what cheating-related behaviors students would be willing to indulge in, or at least condone in their friends, and (2) how "evil" they would judge these behaviors to be under a range of "mitigating" behaviors. A questionnaire was developed wherein five levels of cheating behaviors (advance tips from former students, "accidental" copying, "purposeful" copying, the use of crib-notes, and buying a copy of the test) were alternated in combination with six sets of conditions (everybody else in the course cheating, much lost course-time, a last-minute emergency which precluded any study, low grades combined with a B needed for graduation, the course's simple being "required", and "no special conditions; you just do it") and students' judgments of the relative "evilness" of each situation were elicited via rating scales. The posited levels of evilness were upheld – the five levels *did* receive successively higher ratings of condemnation – but there were no differences found under the various "mitigating" situations, though "very good" and "good" students *were* more condemning of all the behaviors than were "average" students. The authors conclude, then, (a) that these subjects felt that "cheating is cheating, for whatever reasons", and (b) that highly successful students, probably most often the victims of cheating, were generally more condemning of cheating-related behaviors than were less successful students.

A COMPARISON OF VERBAL PERFORMANCE OF FOREIGN AND NATIVE BORN GRADUATE STUDENTS

ELAINE CRABLE
Duke University

Every normal person's mind functions in the same manner when thinking just as every normal person's eyes function in the same manner when seeing. Cultures may cause people to perform these functions differently but the actual mechanics of the function are not different. People in New York City may be thinking and communicating different thoughts than the people in the country of Taiwan or Russia, but all groups are ultimately thinking in the same manner. The thought process is identical, only the languages and particular thoughts differ. Word association study has led many researchers to this conclusion. A specific type of word association study, that of paradigmatic-syntagmatic free word association analysis investigates this question of thought process.

In this type of free word association analysis a high frequency word is usually given to a subject who is to quickly respond verbally with a single word. This response is then classified as a paradigmatic or syntagmatic response according to a set of criteria. Paradigmatic responses have been defined as responses that contrast the stimulus word or are superordinates or coordinates of the stimulus (Dinnan, 1975). Examples of these type of responses would be:

Stimulus	**Paradigmatic Res.**	**Syntagmatic Res.**
mother	father, sister	ugly, country
hot	cold, warm	water, coffee

A substantial amount of studies have demonstrated a close relationship between paradigmatic language response and academic achievement. Research seems to be linked to measure of academic achievement and ability (Dinnan, Bickley, and Williams, 1971; Bickley, 1971; Brosier, 1974). However, research has been done predominately with elementary and high school students and little with the adult and even less research has been done with the foreign student with English as a second language.

The purpose of this study was to assess the thought process as it is reflected in the language performance of foreign graduate students and to compare these to that of native born American graduate students. The test scores of the Paradigmatic Language Inventory (Dinnan, 1973), the Graduate Record Examination (Verbal and Quantitative Ability), the grade-point average, and the results of faculty ratings for the two groups were investigated.

Procedure

The subject included in this study were 88 doctoral-level graduate students enrolled in the University of Georgia graduate school. Forty-four were from five different areas of the world where English was considered a second or third language. The other 44 were from the U.S. with the eastern states most highly represented. Subjects came from many departments at the University of Georgia.

The Paradigmatic Language Inventory (P/S Inventory, Dinnan, 1973) was used as the word association test. The GRE was used to measure academic achievement in the verbal reasoning and quantitative areas. Grade-point average for all subjects was collected and faculty ratings were obtained for each subject.

Each subject was personally contacted in his particular college during a normal school day. Academic information was first collected from the subjects then the subject was given the P/S Inventory. Achievement data for each student was later verified by the researcher who obtained permission to do so from each subject. Two faculty rating sheets were sent out to all advisors and one other instructor familiar with the subject.

Discussion

Student *t*-test was used to measure for any significant differences between the two groups on the variables. The means of the P/S Inventory, the GRE verbal and quantitative, Grade-point average and faculty ratings were compared. Pearson product moment correlations were conducted for investigating relationships between the instruments employed in the study.

The major results of this study were that a significant difference existed between foreign and native born subjects for the P/S Inventory, and the GRE verbal reasoning scores. With the P/S Inventory, the findings tend to support other research such as Penfield's (1969). He concluded that there are two sets of thinking patterns when dealing with a foreign language and felt that a person's hearing a language activates a series of neural connections that forms a set. These patterns or sets are normally kept separate by learners, partly by multisensory cues associated with each language. A person can fluently switch back and forth between languages only if he possesses an equal amount of ability in each language. Unequal ability to activate these multisensory cues seems to be the reason for the lower mean score for the foreign student in the P/S Inventory.

Certain relationships existed between variables for the groups. The two verbal tests, P/S Inventory and GRE verbal section correlated highly with each other, while the quantitative section of the GRE did not relate with any variable in this group. The P/S scores correlated with the Other Faculty Rater, indicating that a faculty member other than a student's advisor may be more objective of a person's ability.

Several recommendations emerge from the results of this study especially concerning the use of the P/S Inventory. An instrument similar to the one employed in this study should be administered to foreign students as part of the admission procedures or during initial stages of the student's program. If the student scores in a syntagmatic manner, he could be tutored both in the area of contrasting and of associating material related to basic language communication, as well as in his specific subject area.

The foreign student needs to be aware that his thinking pattern is syntagmatic in nature. Inability to make relations can cause failure in areas involving verbal ability, both oral and written, in a second language. The foreign born student must understand that shortcomings in a second language do not reflect academic inadequancy. Such misinterpretation could severly effect the student's self-concept. The low GRE verbal reasoning score of the foreign born student yields two further recommendations: 1) admissions officers and other administrators or faculty members involved in admissions should de-emphasize that test score when considering a foreign application; and 2) a special course teaching paradigmatic behavior to the foreign student be designed, thus preparing the foreign student for his academic career in oral written English.

Implications for future research are mainly concerned with studies dealing with thought process investigation. These studies should investigate a person's native language ability as compared to his second language ability. Some suggestions for such research are as follows: 1) a study deaing with the teaching a paradigmatic behaivor to a group of foreign students. McNinch (1972) found paradigmatic behavior could be taught to children. It remains to be seen if this type of verbal behavior can be taught to adults, and, 2) a study should be done to investigate whether a person who exhibits syntagmatic behavior in a second language exhibits predominantly paradigmatic behavior in his native language.

References

Bickley, R. Oral language responses and reading performances in the intermediate grades. *Twentieth Yearbook of the National Reading Conference*, 1971, 11–13.

Brosier, G. The validation of paradigmatic behavior as related to selected variables. Unpublished doctoral dissertation, University of Georgia, 1974.

Dinnan, J. A. The paradigmatic-syntagmatic language inventory. Unpublished test, University of Georgia, 1973.

Dinnan, J. A. Bickley, R., an; Williams, J. Syntagmatic oral responses compared with high and low verbal S.A.T. scores of college freshmen. *Twentieth Yearbook of National Reading Conference*, 1971, 17–19.

Dinnan, J. A. Communication: The logic and meaning of language. Unpublished book, University of Georgia, 1975.

McNinch, G. Paradigmatic language training as a method of increading reading achievement. *Twenty-first Yearbook of the National Reading Conference*, 1972, 100–109.

Penfield, W. Consciousness, memory and man's conditioned reflexes. In Karl H. Pribam (Ed.), *On the Biology of Learning*. New York: Harcourt Brace Jovanovich, 1969, 127–167.

EVALUATION OF READING IMPROVEMENT COURSES: THE CARVER–DARBY CHUNKED READING TEST COMPARED WITH TESTS PREPARED FOR USE WITH THE COURSE MATERIALS

MICHAEL BRADLEY
University of North Carolina – Wilmington

ABSTRACT

The Carver-Darby Chunked Reading Test (CDCRT) is a measure of within-individual change and presents hypothesized categories of mature readers. Data was presented indicating problems of test disagreement when the CDCRT and tests accompanying a specific reading improvement course are both used to evaluate within-individual change. The data was also used to evaluate the usefulness of the categories and accompanying statements concerning the probability of reading training being of benefit to each type of reader.

Though admittedly based on a small sample, the data presented are indicative of some of the problems of the CDCRT categorizations. These problems are extremely evident for the person who wishes to use the CDCRT as a measure of change in a reading improvement course where tests which accompany the course materials are also used. The CDCRT should not be disregarded or discarded in such cases. In fact, the CDCRT is a much better measure of within-individual change than the tests designed for use with the course materials. However, the categorizations and accompanying opinions seem to not be very useful and often a hinderance. Carver and Darby state that the categories are unverified hypothesis and should be treated as such. Data presented here and elsewhere seems to indicate the need for a revamping of many aspects of the CDCRT category system.

THE READING INTERESTS, READING ATTITUDES, AND READING BACKGROUNDS OR PRIVATE–COLLEGE INCOMING FRESHMEN

GEORGE J. CRETILLI
Cardinal Stritch College

KEN L. DULIN
University of Wisconsin

ROBERT CHESTER
University of British Columbia

ABSTRACT

Though private colleges enroll approximately 20% of all those entering college each year for their first time, little research has been done in determining the reading characteristics of this group. This study dealt with these variables in 130 freshmen entering Cardinal Stritch College in Milwaukee, Wisconsin, during the Fall Semester of 1975-76. As part of their registration procedures, these subjects responded to a questionaire dealing with (a) why they read, (b) when they read, (c) for what purposes they read, and (d) various other reading-related factors and topics. This research-group, a relatively "average" one in reading ability but with a "high" attitude towards reading, rank-ordered ten topics to read about from highs of "adventure and romance" and "nature and outdoor life" to lows of "religion and religious life" and "science and technology", and they weighted novels and short-stories well above biographies, poetry, and drama in their preferences for literary genre. Also, they chose soft-bound and hard-bound books over newspapers and magazines for their favorite print-forms, and reported reading much more later in the day than earlier and more at home than in libraries, during travelling, or "in other public places." When asked *why* they read, they reported "for school" and "for leisure" above "for self-education," "for political, civic, and social awareness," and "for religious guidance and inspiration"; and when asked about their own reading competencies they reported high confidence in their abilities to use reference materials but low confidence in their abilities at studying, reviewing for and taking tests, and taking notes while studying. Overall, then, this group was (a) average in reading ability but high in attitude towards reading, (b) made up more of leisure readers than deep, philosophical readers, and (c) better at reference skills than at study-skills in reading.

A SYSTEMATIC APPROACH TO PREDICTING READING PERFORMANCE AT THE SECONDARY LEVEL

CHARLES W. PETERS
Oakland Schools

NATHANIEL A. PETERS
Oakland Schools

At this time there is a paucity of research on the variables predictive of reading performance at the secondary level. Results from studies of teacher evaluation and student self-evaluation do not seem to yield information that is germain to students and teachers in the secondary school. Research investigating the complex interplay between variables having potential predictive capability is required if secondary reading is to progress beyond a reliance on single measures to indicate those who will be able to read and those who will not be able to read content materials. The present study moves beyond the single measure approach to prediction by attempting to analyze the relationship between seemingly related variables.

Method

Decisions made about reading performance at the secondary level are predicated upon many factors such as scores from general reading tests, teacher judgments, and grades, but little attention has been focused upon the relationship between these factors and their ability to predict reading performance. In the present study five variables were analyzed to ascertain just how accurate they were in predicting reading performance. The variables were as follows: (a) general reading performance (Gates-MacGinitie), (b) specific reading performance (teacher-constructed English Reading Test), (c) student self-evaluation (Student Attitude Inventory), (d) teacher assessment of reading ability, and (e) the grades received in previous English courses.

Development of the Content-Reading Test

Since general reading tests do not always reflect the reading skill priorities of teachers, an English Reading Test (ERT) was developed. To ensure that the ERT was predicated upon the reading skills English teachers believed to be necessary for success in their classrooms, they were asked to identify reading skills by using an Essential Skills Checklist. Seven English skills were identified: recognizing main ideas, utilizing contextual analysis, determining realtionships, recognizing details, drawing conclusions, utilizing literary devices, and identifying the proper sequence. Once their reading skill priorities had been determined, teachers were asked to operationally define each skill by placing it in the context of English materials. Through this logical development, content validity was established.

Reliability of the test was calculated by utilizing three different procedures: (a) Kuder-Richardson 20 (.79), (b) Kuder-Richardson 21 (.73), and (c) the split-half method (.82). All three reveal an adequate level of reliability.

Student Attitude Inventory

In order to assess student attitudes toward reading, a student Attitude Inventory was developed. The Student Attitude Inventory (SAI) developed by the authors was subdivided into two parts, a cognitivie component and an affective component. The first section contained items which attempted to assess how the student perceived his own cognitive abilities. Using a Likert scale, students responded to such statements as "When compared to other members of the class, I would tate myself as a better than average reader." The second section of the inventory contained items which attempted to assess how the student felt about reading (affective component). In this section students were asked to respond to such statements as "I like to read materials in this class." The SAI was subdivided in order to ascertain whether there was any difference between either the cognitive or affective component in its ability to predict reading performance. The reliability of the instrument was established by utilizing the split-half method. This procedure produced a reliability coefficient of .79.

Teacher Assessment

One of the most frequent criticisms leveled by classroom teachers against the validity of standardized testing is that it only assesses one aspect of reading performance, the student's ability to respond to an objective test. Since classroom teachers interact with students on a regular basis, it was felt that teacher assessment of reading ability was a variable to consider when attempting to predict reading performance. Teachers were asked to rate each of their students on a scale from one to nine. One represented a very poor reader while nine represented a very good reader. Students were placed into three groups on the basis of the teacher judgment ratings. Those students who received scores from one to three were classified as poor readers. Those students who received scores from four to six were classified as average readers while those students who received scores from seven to nine were classified as good readers.

Grades

In addition to the two measures of reading performance, the student's own assessment, and the classroom teacher's judgment, an additional variable was considered, the grades each student received in English courses. Grades from the two previous years were averaged. On the basis of this average, students were divided into four groups: (a) the A group, (b) the B group, (c) the C group, and (d) the D and F group.

Subjects

The sample consisted of 315 ninth grade students enrolled in English classes. All the subjects were from one middle school. The middle school was choosen, because it represented a range in socio-economic make-up that is reflective of most ninth grade populations.

Each subject participating in the study had the Gates-MacGinitie, Reading Test, Survey E, Form 1 administered to him prior to the ERT and SAI. The raw scores from the vocabulary and comprehension sections of the Gates-MacGinitie were converted to percentiles for the purpose of distinguishing between good, average, and poor readers. Thus, operationally defined good readers were those subjects whose performance placed them between 99th and 67th percentiles, average readers were those subjets whose performance placed them between the 66th and 34th percentiles, and the poor readers were those subjects whose performance placed them between the 33rd and 1st percentiles.

Design

Three statistical procedures were performed. A correlational matrix was computed to ascertain the relationship between not only the five variables (the Gates-MacGinitie, the ERT, the SAI, teacher assessment of reading ability and grades) but also between the two components of the attitude inventory and the seven subskills which comprised the ERT. Once the correlational matrix was computed, an analysis of variance was performed to ascertain whether any difference existed between the Gates-MacGinitie and the four independent variables.

In order to determine the predictability of each of the variables a discriminate analysis procedure was utilized. This process analyzes the relationshp between the independent variable and the dependent variables simultaneously.

Results

To ascertain the relationship between the five variables, a correlational matrix was computed. This computation revealed a low correlation between all the variables except for the ERT (specific reading performance) and the Gates-MacGinitie (general reading performance). The lowest correlation existed between the Gates-MacGinitie and the SAI (affective). The intercorrelational coefficients revealed the extent of the difference that existed between each of the variables:

(1) A moderate relationship existed between Teacher Assessment and Grades (.499).
(2) A low inverse relationshp existed between the SAI (cognitive) and Teacher Assessment (-.406).
(3) A low to moderate inverse relationship existed between SAI (cognitive) and the Gates-MacGinitie (-.454).
(4) A low to moderate relationship existed between the SAI (cognitive) and the ERT (.52).
(5) A low inverse relationship existed between grades and the ERT (-.424).
(6) A low inverse relationship existed between the SAI (affective) and Teacher Assessment (-.341).

The moderate to low correlations seem to suggest that each of the variables appear to be assessing a different element of the reading process. The difference that existed between each of these variables is significant at the .001 level.

Since there was a moderate correlation between the ERT and the SAI, an additional correlational matrix was computed to ascertain whether these two variables were measuring different elements of the reading process. The results not only revealed a low correlation between the SAI (cognitive and affective) and each skill which comprised the ERT, but also a low correlation between each skill contained within the ERT. These low correlational coefficients seem to suggest that first, the SAI assesses an additional dimension of the reading process not accounted for in the ERT; and second, the skills contained within the ERT are separate measures of reading comprehension, a finding not often revealed in most analyses of subskill intercorrelational coefficients. The difference that existed between each variable was significant at the .0001 level.

Each variable appeared to be assessing a different element of the reading process. However, one question still needed further investigation; how useful is this information in predicting reading performance? Three different relationships were investigated to ascertain their predictability: (a) the relationship between the Gates Mac-Ginitie, the seven subskills of the ERT and the cognitive and affective components of the SAI; (b) the relationship between grades, the seven subskills of the ERT, and the cognitive and affective components of the SAI; and (c) the relationship between teacher assessment, the seven subskills of the ERT, and the cognitive and affective components of the SAI. The Gates-MacGinitie, grades, and teacher assessment were the independent variables while seven subskills of the ERT and the cognitive and affective components of theSAI were dependent variables. In order to determine the predictability of each variable, a discriminate analysis procedure was utilized. The relationship between the independent variable and the dependent variables was analyzed simultaneously to ascertain which variables or combination of variables was the most accurate predictor or reading performance.

General Reading Performance

The results revealed that eight of the nine variables provided significantly different information from that accounted for by the Gates-MacGinitie. The Conclusion Subtest of the ERT was the only variable that did not provide separately distinguishable information from that already obtained by administering just the Gates-MacGinitie. The SAI (cognitive) appeared to be the factor which possessed the most variability from general reading performance (Gates-MacGinitie). The difference which existed between each variable was significant at the .001 level.

This combination of eight variables seemed to predict good readers better than eight poor or average readers. This group of variables correctly placed 74.2% of the good readers while only accurately placing 52.9% of the average readers. This process correctly placed 60.3% of all readers.

Grades

Only four out of the nine variables, seemed to provide significantly different information when grades were utilized as a predictor of reading performance. These differences were significant at the .001 level. Again, the SAI (cognitive) seemed to be the factor that possessed the greatest degree of variability.

The students in Group A appear to be the most accurately placed (60%). However, in Groups B and C more students were incorrectly placed than were correctly placed. For example, in Group B approximately 72% of the students were inaccurately placed. The overall accuracy of utilizing grades as a predictor of reading performance was 42.7%. Obviously, grades, the ERT, and the SAI were not as predictive of reading performance as the Gates-MacGinitie, the ERT and the SAI.

Teacher Assessment

Only five of the nine variables (details, conclusions, sequence, SAI [cognitive], and contextual analysis) seemed to provide significantly different information when Teacher Assessment was utilized as a predictor of reading performance. These differences were significant at the < .001 level. The Detail Subtest was the factor that possessed the greatest degree of variability. This process (Teacher Assessment, the ERT and the SAI) seems to place more accurately those students who comprised the good and average reader groups. However, in the group which contained the average readers this process placed only 34% of the students correctly. The overall ability of this process to accurately place all readers was 53.9%.

Discussion

As the earlier research seems to suggest, each of the five variables (Gates-MacGinitie, the ERT, Teacher Assessment, the SAI, and Grades) investigated in the present study appear to be measuring different elements of the reading process. The data derived from the correlational matrix and that generated by the discriminate analysis procedure verify these conclusions. While the difference between each variable was significant at the < .001 level, not all the information provided in the analysis can be considered uniquely different.

The information obtained from the affective component of the SAI did not seem to increase predictive accuracy. Two out of the three procedures utilized to predict reading performance did not find this variable to be significantly different from information already obtained. While one's attitude toward reading is a significant factor to consider, it does not appear to contribute uniquely different information when the concern is predicting reading performance in secondary classrooms.

A second finding the research seems to confirm is that grades, as an isolated variable or in combination with the ERT or the SAI, proved to be the least predictive of reading performance. This process was only accurate in placing 42.3% of the students. Even when grades were considered in conjunction with teacher assessment, the salutary information derived from the combination of variables seemed to be limited. While grades are an important factor to consider, the results of the study seem to suggest that they should be utilized with some discretion.

The single variable that seemed to be the most predictive of reading performance when considered in conjunction with general and specific reading performance was the cognitive component of the SAI. The relationship between these variables seemed to suggest that a diagnostic process which seeks to incorporate the student's assessment of his abilities provides valuable information. As classroom teachers attempt to gain more accurate information in regard to reading performance, the student's perceptions of his performance not only provide distinctly different information, but also information that increases the accuracy with which one can classify students according to reading performance. As the research related to student evaluation seems to suggest, the vast majority of it has tended to focus upon attitudes toward reading without much concern for the cognitive dimension. As was indicated earlier, the affective domain is an important variable. But it seems to have very little significance when attempting to predict reading performance. Therefore, if predicting reading performance or accurately identifying reading abilities is an important consideration, then the cognitive component of the self-evaluation process needs to be incorporated into the assessment process. By using general reading performance, in conjunction with specific reading performance and the cognitive component of the SAI, it is possible to accurately predict 60% of the students; according to present standards, this is a significant number to predict accurately.

When Teacher Assessment was considered in conjunction with the ERT and the SAI as predictors of reading performance, this process was still not as accurate in the placement of students as the first combination of variables (Gates-MacGinitie, the ERT, and the SAI). While there was a significant relationship between Teacher Assessment and each of the above variables, it still did not seem to provide distinctly different information. The literature investigating the relationship between Teacher Assessment and reading performance has indicated that secondary teachers have difficulty identifying students with reading problems. The results of the present research seem to confirm those findings. Teacher Assessment, while not proving to be the most accurate of the three procedures, nevertheless, provides information that should be incorporated into the diagnostic process. As long as this information is viewed in its proper perspective, it can be useful.

Two additional results were revealed in the study. First, specific reading performance and general reading performance seem to provide separately distinguishable information. The combined information from both instruments seems to improve the predictability of reading performance. This seems to support earlier research (Peters, Peters, & Kaufman, 1975; Peters, 1975) investigating the relationship between general and specific reading performance; and second, the subskills which comprised the ERT appear to be assessing separate segments of the reading process. This seems to support the research which suggests that comprehension is a composite of separately identifiable skills.

References

Peters, Charles W.; Peters, Nathaniel A.; & Kaufman, B. Darwin. A comparative analysis of reading comprehension in four content areas in the Twenty-Fourth yearbook of the National Reading Conference, 1975 (In press).

Peters, Nathaniel A.; & Dulin, Kenneth L., A comparison of teacher-judgment and criterion-evaluation of elementary school pupil's reading abilities. Wisconsin Educational Research Association, Madison, February 2, 1975.

Test References

GATES–MCGINITIE READING TEST, Survey E., Form 1. Arthur L. Gates and Walter H. MacGinitie. New York: Teachers College Press, Columbia University, 1970.

COGNITIVE AND PERSONALITY PREDICTORS OF LISTENING ABILITY IN COLLEGE STUDENTS

JANE GLOTFELTY RHOADS
Wichita State University

PATRICIA ELLIS
Wichita State University

A.B. SWENEY
Wichita State University

ABSTRACT

The purpose of this study was to discover what factors determine or contribute to a student's listening ability.

The subjects for the study were 163 students enrolled in Reading Improvement Techniques, a two-hour, credit course at Wichita State University. The sample contained all grade levels with a close balance between sexes. The tests used in the study were the *Diagnostic Reading Test – Form A & B: Brown-Carlsen Listening Comprehension Test – Form Bm; SRA Nonverbal Form – Form AH; Study Habits Inventory* by Gilbert Wrenn: the *Hours of Reading Inventory; Reading Skills* by Baker, Chapters 1 and 16; and Cattell's *16PF Personality Inventory – Form C.*

All of the tests were administered before the course work was begun. After scoring the instruments by the standard procedures, the data were correlated with Pearson Product moment coefficients to examine the bivariate relationships. Since large intercorrelations were found among the interdependent variables, an orthogonal Varimax factor analysis was conducted omitting the listening variable to simplify the possible sources of contribution. Factor scores for the ten factors obtained were correlated with the listening variables to analyze the sources of factor variance. The three factors which accounted for 50% of the measured variance of the total scores were identified as Reading Comprehension, Reading Speed, and Freedom from Preoccupation.

Although many of the bivariate correlations showed significant relationships with the listening scores, they overlap and present a confusing picture of the variance actually contributed. The correlations with the factor scores indicate that around fifty per cent of the variance for the listening behaviors is accounted for by the independent factor

scores in this study. Whereas speed and concentration factors are readily explained, the lesser correlations of Preoccupation, Female Sensitivity, and Control provide some difficulty. Are females slightly more prone to "send" than "receive?" Do communication channels get overloaded and inhibit listening to new material? What is the nature of Control and what implication does this hold for our concept of listening as an exercise in passivity? Does the negative correlation with conscientiousness mean that the rebel hears better that which he denies than the conscientious individual does that which he is preprogrammed to accept?

Brown-Carlsen Scale I is additionally affected by non-cognitive factors which include Female Sensitivity, Control, and Aging Crystalization. Brown-Carlsen Scale II is positively related to Comprehension, Speed, Lack of Preoccupation, and Female Involvement, but negatively related to Anxiety and Control.

THE NEED FOR STUDY SKILLS AS PERCEIVED BY COMMUNITY COLLEGE STUDENTS

DORINDA ANN MCCLELLAN
Hillsborough Community College

LESLIE MCCLELLAN
University of South Florida

ABSTRACT

The purpose of this study was to determine the need for assistance in study habits, skills. and related learning problems as perceived by community college student. Research findings were primarily concerned with identification of five major reading-study problems areas by the distribution of frequency of responses indicated by student-perceived need for assistance in the areas (study management, course related skills, auxilliary course skills, attitutdes, interests, and habits, and physiological aspects). The Survey of Reading/Study Efficiency designed by Frank L. Christ and published by the Science Research Associates was administered to 246 community college students. The subjects, both male and female, ranged in age from 18 years to 65 years with diverse backgrounds of work experience, career aspirations, and educational potential.

The perception of the need for study skills of community college students is as varied and unique as the student's interests, abilities, and educational background. The students in this population indicated difficulty with spelling, technical vocabulary, comprehension, library research, marking and underlining, attitude toward reading, and scheduling. The areas which seemed to cause the least concern for the students was flexibility in reading, choice of careers, attitudes toward school, and classroom testing. Students did not find adequate time for studying yet they provided for ample recreational and relaxation periods. Making notes and listening to lectures as well as summarizing the notes and lectures caused problems for these students although only 20% of the population reported trouble with test taking. About 50% of this population cannot and does not utilize the library and its resources, does not utilize auxiliary course materials, and the same number experienced difficulty with technical vocabulary. A majority of the students do not read books and only 33% of them read daily

newspapers. Only 18% had not made a career choice and 92% were not considering changing thier course major. One third of the students did not like to study. An instrument of this type must be viewed with caution because of the subjective nature of the data obtained. Additional testing and/or observations must be utilized in order to determine the accuracy of student self-assessment.

RELATIONSHIP BETWEEN PERFORMANCE AND INTEREST IN ADULT EDUCATION

THOMAS P. FITZGERALD
State Education Department – New York

ABSTRACT

The literature available concerning curriculum development for adult education programs stress the need for relevance, readability and interest as critical factors with instructional material. The term interest has been studied under various definitions but such reports have not provided consistent data linking higher interest in material with superior performance. Most studies dealing with the notion of interest have used elementary or secondary subjects and have failed to provide a control of the level of material difficulty or the reading achievement of the subjects.

The present study focused on the possible relationships between performance and expressed interest in individual passages, and expressed interest by categories of passages. One-hundred and eighty adult education students from three instructional centers in New York State were administered passages on one of three levels of difficulty. Following each passage, the students answer questions in a multiple choice format which evaluated the subjects recognition of stated details from the passage. After reading all of the selections the students rated each selection as either one of the most interesting, one of moderate interest, or one of the least interesting. The results of this forced choice procedure were correlated with the students performance on the 25 questions for each level of material difficulty. The data on expressed interest for passages were re-analyzed using the five categories of the material used in this study: history, people, environment, body function, and animal life. The performance data were then correlated with the results of the interest factor by categories.

An analysis of the data indicated negative correlations between passage interest and performance at the easy (-.48) and difficult (-.17) material levels. However, with moderate material, a high positive correlation (.83) was generated. When the data for the high reader group (6.7 to 8.0 grade equivalent) and low reader group (4.8 to 6.2 grade equivalent) were analyzed separately, comparable correlations were produced by both groups with easy (-.34 and -.46) and moderate (.58 and .41) material. The high group achieved a negative correlation

(-.21) between performance and interest with *the difficult materials* while the low group achieved a positive correlation (.35).

Since consistent results were not produced analyzing material by levels of difficulty, a further review of the data by categories was completed. Positive correlations were indicated between performance and interest by categories for all subjects (.60) as well as the high reader group (.60) and low reader group (.68). When all subjects ranked passages on an interest factor, the results described the categories used in this study from most interesting to least interesting as follows: people, functions of the body, animal life, history and environment. The performance of all subjects resulted in a similar ranking of categories but with animal life and environment reversing positions in this ranking by performance.

The results tend to indicate that the readability levels, as determined by the formulas used in this study, may not feature all the factors which influence the comprehensibility of material for adults. The moderate material which was closest to the listed reading achievement levels of these readers produced the highest performance scores. It also appears that low reader groups may be more sensitive to interest by category of material than high reader groups. The factors of motivation and self-image may be operating more strongly for low readers since they may tend to react negatively with material without intrinsic appeal. Further work is needed since the test conditions of this study may have influenced performance over an interest factor. The operation of motivation, material relevancy and the material interest may be more critical factors with instruction or when a choice of material is offered than under test conditions which operated in this study.

AN INSTRUMENT DESIGNED TO IDENTIFY SPECIFIC STRENGTHS AND WEAK NESSES IN READING INSTRUCTION

BOB W. JERROLDS
University of Georgia

SIDNEY E. BETON
North Georgia College

JUDY BAZEMORE
Rand-Spartenburg, S.C.

Development of the Instrument

As a result of dissatisfaction with subjective supervisory judgments, paper and pencil tests and Likert-like rating scales for the evaluation of reading programs, the Atlanta Public Schools initiated a project to develop a behaviorally oriented, objective instrument that would aid in judging teacher competency in reading instruction.

It was decided to build the instrument on two data sources: observation of classroom teachers who were teaching reading and the writings of authorities in the reading field. In the process of collecting data from the first group 43 hours of observations were made of the teaching of reading in 28 teachers' classrooms. Reading supervisors had been asked to indicate which of these teachers had classrooms in which the children were effectively learning to read and those who had classrooms where the children were not effectively learning to read. Systematic observations were made of these teachers noting factors that could be actually observed, factors that could be measured, practices and materials that appeared to be effective and those that appeared to be ineffective.

A great number of observable, promising practices began to cluster in the records made of observations of effective teachers. Such practices as having children assigned to levels of materials that they could handle, fluid grouping and reading aloud to the children were rapidly evident with these teachers. Particularly gratifying was the fact that there were non-verbal responses so clearly evident as to indicate that they could be considered in a behaviorally oriented observation instrument.

With all honesty it must be admitted that apparently highly effective teaching was being done by some teachers using archaic and bizarre practices. After due consideration it was decided that most of these practices could not be effectively used by anyone except the particular teacher involved and could not be worked into a practical instrument.

Observations of ineffective teachers were highly edifying but contributed no items to the instrument in its final form. A great many manifest behaviors were recorded that seemed to be related to poor teaching. It was later discovered, however, that when the positives and negatives were combined into one instrument, the administration and interpretation of that instrument became hopelessly confusing.

All the negatives or factors identified with poor teaching were dropped, except in a few instances where they were placed as qualifiers on the positive factors. For example in item 22 we have the positive that "... the teacher called upon each child to read at least once during the oral reading activity." but with this qualifier, "Without resorting to 'round robin reading' ...."

The second data source was the authorities in the field of reading. Fifteen basic textbooks on the teaching of reading in the elementary and high school were consulted for those elements that the reading people had a reasonable amount of agreement on as to what ought to constitute a reading program.

The authors of the instrument spent several days translating the original observations of teachers and suggestions of the authorities into questions, the answers to which could be observed or assessed according to behavioral criteria.

The resulting first draft of the instrument was submitted to a group of reading supervisors whose full time jobs were supervising reading instruction in the Atlanta Public Schools.

They were asked to examine the draft with the following questions:

1. Can these things be observed?
2. Are these things truly manifestations of effective teaching of reading?
3. How long do you think it will take to administer the instrument?
4. How can the instrument be made more practical?
5. How can the wording of the instrument be improved so as to make it more objective?

On the basis of this examination and feedback the instrument was revised. It was decided that some factors would take so long to observe as to make the process impractical. Several of these problem items were converted into questions to be directly asked of the teacher.

Five items were dropped entirely because they were deemed to be repetitious. A great number of items were revised. Two were added because the group felt them to be important and untapped areas.

The second revision of the instrument was given to the reading supervisors one week later. After discussion of administration procedures, each of these supervisors was asked to use the instrument in observations of two teachers and to make notes that would be used for another revision.

As a result of feedback from this field testing a number of minor changes were made. The only major concern was that the time required for administration was still found to be intolerable. Therefore, items were taken out of format where all oral reading questions were together, all readiness questions together, etc. and the format changed to maximize efficiency of marking.

The third revision was submitted to a group of principals and to a group representative of the teachers' organizations. Both groups expressed opinions that it was good to have an instrument like this one, that it was better than evaluative approaches taken in the past and there were grave dangers that such an instrument could be misused. Few actual changes were made in the instrument itself in this revision.

The fourth revision was submitted informally to a number of professors of reading education with requests for reactions. These reactions were considered by the authors and a few changes made. This revision was also submitted to 18 doctoral level students in reading education at the University of Georgia who were enrolled in a seminar. They were asked to make any suggestions they felt to be pertinent and to respond to the following questions in writing:

1. Does this instrument cover, in a reasonable comprehensive fashion, the field of reading instruction?
2. Are there areas that are over-emphasized?
3. Are there areas that should be added?
4. Are there items that are redundant?
5. In your estimation would a high score on this instrument have a high positive correlation with good teaching of reading? In your estimation would a low score on this instrument have a high positive correlation with poor teaching of reading?
6. What items do you feel demand too much?
7. What items do you feel demand too little?
8. What are your suggestions for making the instrument more efficent?
9. What uses could be made of this instrument?
10. What are your suggestions for scoring?
11. What are your suggestions for improving the directions?

The present form of the instrument is the result of a revision made following a number of suggestions from the doctoral level students.

DOCTOR Description

The first section of the instrument consists of 29 items that are to be marked by the observer during a reading lesson or a reading activity in a content area. Examples of such items are: (a) at least **one** thing other than the story to be read was used as a part of the readiness activity, and (b) the questions asked after silent reading were in line with the purposes set for silent reading.

The secod section consists of 6 items. The observer randomly selects, one at a time, five students. Each student is asked to read to the observer from a library book that he has checked out and from a book selected at random by the observer from among the student's assigned reader and content area books. Examples of items in this section are: (a) four out of five randomly selected students read from their assigned texts with at least 95% accuracy of word attack, and (b) at least three out of five students had a book checked out of the school or public library.

The third section consists of 19 items that the observer marks as a result of an interview with the teacher in which the observer asks specific questions and asks to see materials relative to the teacher's reading instruction. Examples of the items in the third section are: (a) upon request the teacher identifies all those in the classroom who do not like to read, and (b) upon request the teacher showed the observer two pieces of teacher prepared materials he uses to teach context clues, structural analysis or phonic analysis.

DOCTOR Reliability and Criterion Validity Study

Evidence concering reliability and criterion-related validity was obtained using 25 teachers in the Butts County school system of Georgia. Butts County is fairly representative of the rural South. The overall socio-economic level is low to middle.

Twelve upper elementary and 13 primary school teachers were randomly selected from the total of 51 elementary teachers in the system. The 25 teachers were then randomly assigned to one of two reading supervisors in the county. The reading supervisors visited the classrooms and administered the DOCTOR to the teachers. Each item on the instrument was scored as: −1 for "No," O for "Not Applicable" and +1 for "Yes." Total scores on the DOCTOR and scores on each of the three sections were calculated for 24 teachers (one teacher was evasive and not cooperative; thus, she was eliminated from the study).

The estimated reliability of DOCOTR, using the Spearman-Brown formula, was .83. Criterion measures used were (a) Artley and Hardin's Inventory of Teacher Knowledge of Reading (1971 Edition) administered, but not examined or scored, by the reading supervisors and (b) principal appraisals of the teachers on a 5-point Likert-type scale completed without the principals' knowledge of the DOCTOR results. A step-wise multiple correlational analysis was carried out for each criterion with scores on three sections of the DOCTOR as predictors. Correlations based on the total scores of the DOCTOR and the two criterion measures were also calculated.

The correlations for the DOCTOR section scores, principal ratings, and scores on the Inventory of Teacher Knowledge of Reading ranged from −.09 to .30. The Artley and Hardin instrument correlated .18

with principal ratings which is comparable to the three correlations of the sections (Directed Reading Activity, Reading of Students, Teacher Interview) of DOCTOR with principal ratings: –.09, .23, and .16, respectively. None of the correlations were significant at the .05 level. However, the sizes of the intercorrelations for the three sections of the DOCTOR indicate that the sections are relatively independent of each other.

In the first step-wise regression analysis the three DOCTOR section predictor scores were entered in the following order: section 1–Directed Reading Activity, section 2–Reading of Students, and section 3–Teacher Interview. The multiple-Rs obtained using the three sections to predict principal ratings were .23, .28, and .30. None of the F values were significant at the .05 level. Only 9% of the variance could be predicted using the three sections as predictors.

In the second step-wise regression analysis the three DOCTOR section predictor scores were entered in the following order: section 3–Teacher Interview, section 1–Directed Reading Activity, and section 2–Reading of Students. The multiple–R obtained using the three DOCTOR sectors to predict scores on the Inventory of Teacher Knowledge of Reading was .36. Neither this value nor any of the other multiple-Rs was significant at the .05 level.

The correlation between total scores of the DOCTOR and principal ratings was .04, and the correlation between total scores on the DOCTOR and Inventory of Teacher Knowledge of Reading was .29. Neither of these correlations was significant at the .05 level.

In summary there was no section or combination of sections of the DOCTOR that contributed significantly at the .05 level to the prediction of either principal ratings or scores on the Inventory of Teacher Knowledge of Reading. The results then indicated that neither section scores or total scores on the DOCTOR were accurate predictors of principal ratings or scores on the Artley and Hardin instrument. These findings are similar to those of Scott (1973) who found no significant section or sections of a student appraisal instrument in predicting college supervisor ratings of student teachers in high school English classes.

Scott stated, "this lack of realtionship was not altogether unexpected. College supervisors had limited first-hand knowledge of what went on in the classroom since each supervisor made five visits or less to the student teacher's class during the quarter . . ." (p. 5). A parrallel might be drawn between the supervisors in Scott's study and the principals in this study.

Artley (1975) expresses a similar caution in the interpretation of the scores from the Inventory. " . . . it should be kept in mind that the instrument is a measure of an individual's understanding of reading in terms of the items included. In no way does it indicate the extent the teacher uses these understandings in the classroom." In the present study no significant relationship was found between what a teacher knows about the teaching of reading and the observable practices that take place in that teacher's classroom.

Implications for Use of DOCTOR

There are several limitations that should be kept in mind in considering use of this instrument:

1. The administration of the instrument is very time consuming and oten requires a repeated visit. The least reported time was 30 minutes; the greatest reported time was 2 hours.
2. If a particular teacher wants to prepare just for the instrument and achieve a spuriously high assessment, it can be done.
3. Totally individualized and totally programmed reading instructional programs cannot be evaluated by means of this instrument.
4. If misused the instrument could be very destructive rather than helpful.

Strengths of the instrument are:

1. Its design and revisions were carried out with more rigor than the design and revision of many other instruments used to assess teacher effectiveness.
2. It is more in line with today's emphasis on competency based instructional procedures.
3. It is more objective than many teacher evaluative instruments and procedures.
4. It samples from the most important dimensions of effective reading instruction.
5. It provides for the identification of specific areas of strengths and weaknesses in the reading instructional program.
6. Most of the items on the instrument could be used by the teacher in an analysis of his own teaching.

The instrument may be obtained from the author and reproduced with the written consent of the authors and the Atlanta Public Schools.

References

Artley, A. S. Personal Communication, October 20, 1975.

Scott, O. *The measurement and use of student perceptions to improve college instruction*. Paper presented at the annual meeting of the Natioal Council of Measurement in Education, New Orleans, Louisiana, 1973.

ASCERTAINING KNOWLEDGE OF READING WITH THE ARTLEY–HARDIN INVENTORY

KARL KOENKE
University of Illinois

In a recent publication, Kingston, Brosier, and Hsu (1975) reported the first validation of the *Inventory of Teacher Knowledge of Reading* (ITKR) by Artley and Hardin (1971).

However, since the subjects used by Kingston, et al, were either enrolled in classes at one large university or associated with it in some fashion, it behooves teacher educators and reading specialists from other institutions to inaugurate their own validation studies of the ITKR. This then is a report of a validation of the ITKR using students enrolled in Elementary Education at the University of Illinois and experienced elementary school teachers from central Illinois. In addition, consideration is given to identification of ITKR item categories since the results of the factor analysis conducted by Kingston, et al, did not yield sub-scales. Finally, consideration is given to the selection of items which might be used in a short-form of the ITKR.

Procedure

During the 1974–75 academic year, the *Inventory of Teacher Knowledge of Reading* was administered to 60 University of Illinois freshmen, 60 juniors, 60 seniors, and 60 elementary school teachers currently employed in the school districts of central Illinois. An attempt was made to randomly select the undergraduates from the total at each of the three levels; however, not all of the first 60 students would or could come to take the test. In all instances where this occurred, alternates were randomly chosen from the remaining students at that particular level.

The undergraduates were all female and enrolled in the elementary education curriculum at the University of Illinois, Urbana-Champaign. The 60 freshmen were selected from the 219 entering freshmen and tested during the orientation and registration period. The 60 juniors were selected from among the 175 third-year students then enrolled. Testing was conducted as the juniors completed the integrated general elementary methods course and observation of teacher practice in the schools. The 60 seniors were selected from among the 156 fourth-year students then enrolled. They were tested six weeks after the semester started; that is, during the time that they were concurrently enrolled in three eight-week methods courses and just prior to leaving for a half semester supervised teaching experience. Two of the methods courses were required: Fundamentals of Reading Instruction and Language Arts in the Elementary School.

It is, therefore, possible to think of the freshmen as without a college education, the juniors as having both two years of a liberal arts education and an introduction to the elementary schools, while the seniors have most of the methods course-work, including reading, but lack experience.

The experienced teachers, however, were not randomly selected although they represent those teachers who were entering the Master of Education program and enrolled in their first graduate reading course. Half of the sample came from an advanced off-campus developmental reading class and the other half from several sections of the first graduate course in remedial reading. Other characteristics of the experienced teachers are as follows: (a) they were all female; (b) they had 2 to 30 years experience; and (c) they had taken their undergraduate work at other institutions. The few experienced teachers in the classes who had taken undergraduate degrees at the University of Illinois were not included in the study since the usual M.Ed. student at the University of Illinois was not at the University of Illinois as an undergraduate.

Results of Statistical Analysis

Table 1 presents for each of the four groups range, median, mean, standard deviation, and both KR 20 and KR 21 reliability coefficents. It is apparent that test scores increase as education increases, that medians and means are quite similar so the distributions are not skewed, and that reliability coefficients are generally in the *.70s*.

A comparison of the scores of the four groups using an analysis of variance produced significant $F = 74.7$. The Newman-Keuls method for the use of the modified q statistic was used to prove the nature of the group differences after the significant overall F was obtained. The results indicate that in all but one instance each set of scores differs from all others. The one exception is as might be expected from inspection is that seniors' and experienced teachers' scores on the ITKR do not differ significantly. It is interesting to note that Kingston, et al, (1975) also found that the scores of their undergraduates ($\overline{X} = 54.45$) and the scores of their elementary school teachers ($\overline{X} = 54.11$) did not differ.

Another result of this study which is similar to that of Kingston, et al, is the level of the KR 20 reliability coefficients. They calculated a .74 for students with reading courses and a .79 for students without. The seniors with a reading course in this study had a .73 while both the freshmen and juniors without reading courses had a .78. It is only the reliability coefficients of the experienced teachers which differ (.82 to .91), and the smaller standard deviation (9.6 to 14.3) of the sample in the present study might lead one to believe that an attenuated range of scores earned by the experienced teachers in the present study contributed to the lower reliability coefficient.

From the selected item analysis data presented in Table 1, it is apparent that three items functioned similarly across all groups.

Table 1
Items Passed by 90% or Failed By More Than 30%

Group	N	Items passed by 90 percent or more	Items failed by 30 percent or more
Freshmen	60	3	13, 26, 27, 31, 38, 40, 47, 48, 50, 58, 59, 71, 75, 78, 79, 80, 84, 86, 88, 91, 93, 95
Juniors	60	3, 11, 17, 20, 32, 52, 82	26, 27, 31, 37, 47, 60, 69, 71, 78, 80, 88, 95
Seniors	60	1, 3. 4, 6, 8, 9, 11, 12, 14, 17, 24, 32, 34, 52, 54, 57, 67, 72, 73, 82	37, 47, 80, 95
Experienced Teachers	60	1, 2, 3, 5, 11, 14, 21, 32, 51, 67, 82	47, 69, 88, 95
TOTAL	240	3	47, 95

Item 3 was known by only 10% of each group and two items, 47 and 95, were answered incorrectly by more than 30% of each group. The findings of Kingston, et al, agree but items 71, 80, and 88 are also considered difficult.

As might be expected, especially since the analysis of variance was significant, the freshmen have the greatest number of items which caused difficulty among the test-takers and the fewest number of items passed by 90% of the takers. On the other hand, the seniors and the experienced teachers, although not showing a difference in the number of items failed by 30% or more, do differ in the number of items passed by 90% or more of the takers, the seniors succeeding on 20 items while the experienced teachers succeeded on but 11 items. A test of the significance of the difference between the two percentages, 20/95 = .21

and 11/95 = .115, led to the calculation of $t = 1.81$. A significant difference is not claimed, however, since a t value of 1.96 is needed for significance at the .05 level.

Results of the Logical Analysis

Although Artley and Hardin write that their IKRT contains items which sample knowledge in Reading Readiness, Word Perception, Comprehension and Critical Reading, Evaluation, Diagnosis, Correction, Goals of Instruction, Differentiated Reading Instruction, and Silent and Oral Reading; they do not list the items in each knowledge category. This is precisely the information a user needs, however, to make some judgments about the IKRT.

The items of the IKRT were therefore assigned to one of the knowledge categories identified by Artley and Hardin. The results are shown in Table 2. It should be noted that three reading specialists independently categorized the items, and when differences arose a majority opinion was used to assign the item to a category.

Table 2
Items in Logically–Derived Reading–Knowledge Categories

Reading Readiness	1, 9, 11, 19, 23, 34, 40, 57, 63, 66, 68, 70, 66, 68, 70, 76, 83, 89, 95
Word Perception	2, 3, 6, 8, 12, 14, 17, 27, 39, 44, 50, 61, 62, 82, 86, 88
Comprehension & Critical Reading	10, 15, 22, 26, 30, 33, 41, 42, 46, 52, 65, 69, 77, 87, 90, 94
Differentiated Reading Instruction	24, 37, 45, 48, 49, 55, 67, 71, 73, 75, 79, 81
Evaluation, Diagnosis, Correct	4, 21, 25, 32, 38, 43, 51, 56, 20, 28, 64, 74 80, 92, 16, 18
Goals of Instruction	5, 7, 13, 35, 53, 58, 60, 91, 93
Silent and Oral Reading	29, 31, 36, 47, 54, 59, 72, 78, 84, 85

It is apparent after studying Table 3 that the IKRT is heavily weighted in favor of Reading Readiness, Word Perception, and Comprehension-Critical Reading. With 16 items per cateogry, these three categories account for 48 items or 50% of the test. In contrast, Evaluation, Diagnosis and Correction items were displayed jointly in Table 3 because they were assigned respectively eight, six and two items.

Even though Kingston, et al, found so many components in their factor analysis of the IKRT that they despaired labeling any sub-scales, an attempt was made to establish the reliabilities of the knowledge categories using the seniors' data. The KR 20 reliability coefficients for the knowledge categories ranged from .00 to .475; that is, so low as to cause doubt as to the usefulness of the knowledge categories as independent scales.

Since the lack of reliability of the sub-scales may be due in part to the low reliabilities of the items within the sub-scales, an inspection of the item analysis data was undertaken to ascertain which items had low Point Biserial Correlation Coefficients. Low was defined as .25 and below, mediocre as .26-.39 and high as .40 and above. It should be kept in mind that .80 is the highest Point Biserial Coefficient obtainable (Description and Interpretation of MERMAC System: Item Analysis).

The results of the inspection of the item analysis data can be summarized as follows: only 29 items (30.6%) meet or exceed the .25 cut-off, and only three of the knowledge categories have five or more statistically acceptable items, while none of the Diagnosis items and only 3 of the 16 (18.75%) Reading Readiness items meet the criterion.

The final logical analysis of the data to be undertaken was a comparison and contrast of the item analysis data of the juniors, the seniors, and the experienced teachers. The objective was to find items which might be used for a short screening test. The only items which were eligible were those which had a Point Biserial Coefficient above .25 at the junior level and at either the senior, or experienced teacher levels. In addition, the proportion of correct responses had to rise from junior to senior or experienced teacher. Twenty items met these criteria.

Conclusions and Implications

The conclusions and implications of this study are as follows:

1. The IKRT is a valid measure of teacher knowledge of reading in that it discriminates among groups of females with varying degrees of professional training.
2. The reliability estimates were generally in the .70s for this sample. That is considered low for published tests.
3. The low reliability may be due in part to the low item correlations. As Groff (1973) has suggested in his review of the IKRT differences of professional opinion regarding the "truth" may in part be the cause of this condition.
4. Whether or not more instruction is needed for in-service or preservice teachers is a question not answered by this study.

However, the fact that both Kingston, et al, (1975) and the present study found scores of students with reading methods and scores of experienced elementary school teachers to be similar should be the basis for reflection among reading specialists.

5. Although IKRT items were assigned to the knowledge categories identified by Artley and Hardin, the low sub-scale reliability estimates seem to preclude the use of independent sub-scales for diagnostic purposes.
6. The fact that 50% of the IKRT items were assigned to but three knowledge categories: Reading Readiness, Word Perception, and Comprehension-Critical Reading, should alert the potential user to the possible mismatch between the emphasis placed on an objective in the reading course and the emphasis placed upon that objective in the IKRT.
7. If the need for a short-form of the IKRT were felt the present data would support the choice of certain items.

References

Artley, A. Sterl, and Hardin, Veralee B. *Inventory of Teacher Knowledge of Reading*. Columbia, Mo.: Lucas Brothers, Publishers, 1971.

"Description and Interpretation of the MERMAC System: Item Analysis." Mimeo. University of Illinois. Undated. 9 pages.

Groff, Patrick. "Reviews: Artley, A. S. and Hardin, V. B.," *Inventory of Teacher Knowledge of Reading. Reading World,* Volume 12, Number 4, May, 1973, 296-98.

Kingston, A. J., Brosier, G. F., and Hsu, Y. "The Inventory of Teacher Knowledge of Reading—a Validation." *The Reading Teacher,* Volume 29, Number 2, November, 1975, 133—136.

FUNCTIONAL LITERACY OF TEACHER TRAINEES

EMILIE P. SULLIVAN
University of Texas–El Paso

Functional literacy is not just the attainment of a particular grade-level of reading efficiency, but must be defined relative to an individual's job needs. Functional literacy may be considered as having attained that level of reading skill at which reading performance does not hinder job success for contribute to differentiating among job-seekers (Sticht, 1972).

The problem of this research concerns the functional literacy skills of elementary school teacher trainees. A major underlying premise of this paper is that the teaching of reading to elementary school children is a basic task of the classroom teacher. To accomplish this task, the majority of teachers use basal reading programs which include an accompanying teacher's guide. If the teacher is to be able to use the guide and, hence, the entire reading program as designed, she must be able to read the teacher's guide with understanding. A review of the literature indicates a paucity of research on the question of teacher trainee literacy skills. In an earlier study Geeslin and York (1971) found minimum functional literacy skills among inservice teachers when using the *Nelson–Denny Reading Test* (NDRT) as a measurement instrument. The present study measures functional literacy skills in terms of elementary teacher trainees' performance on cloze passages which have been abstracted from teacher's manuals of four basal reading series programs. The following three questions were investigated: (a) what is the relative reading difficulty of the sample teacher guides, (b) what proportion of trainees are functionally literate, and (c) is there a relationship between professional training and functional literacy levels.

Method

Subjects

Subjects of this study were 57 elementary education majors in their junior and senior years of university training who were currently enrolled in education reading courses. An analysis of the trainees' NDRT performance indicated the following: (a) the lowest total reading score was 10.3 grade-level, (b) the median test score was 13.3 grade-level, and (c) 48% of the trainees scored above the test norms of 14.0 grade-level.

Instrument

The passages used to measure functional literacy were abstracted from the teacher guides of four state-adopted basal reading programs. The reading programs utilized were the Ginn 360 Series (G), the Bookmark Reading Program (BM), the Houghton Mifflin Readers (HM), and The Macmillan Reading Program (M). Each cloze test passage was approximately 200 words in length and contained an introductory statement of the approach or philosophy of the particular reading program. A fifth word deletion system was employed. Trainees were given in random order four cloze passages to complete. No time limits were established and work was accomplished within the classroom setting as part of the regular coursework activity. Completed passages were scored on the basis of exact replacement of missing terms and percentage of correct completion scores were computed.

Data Analysis

For purposes of data analysis the trainees in this study were divided into three groups according to the number of reading courses completed. Group 1 consisted of college juniors who were enrolled in their first reading course; Group 2 consisted of college juniors in their second reading course; Group 3 trainees were college seniors enrolled in their third reading course. For each group cloze mean scores and percentages were computed. Each cloze passage was also tested using the *Flesch Reading Base Formula* (Flesch, 1951) to determine the reading ease score. These data were then related to the trainees NDRT scores and cloze test performance.

Results

The sampled teacher guides displayed a wide range of relative reading difficulty when accessed in terms of teacher trainee performance. Table 1 shows the mean cloze raw scores for each group of trainees on each basal reading series passage. Mean raw scores for all trainees are also given. The mean cloze raw score for all trainees on all passages was 15.4 correct out of a possible 41 or 38% of correct completion.

Table 2 shows the percentage of subjects who made a cloze score of approximately 40% or higher on each passage. According to Rankin and Culhane (1969) a cloze score of this level is equivalent to the conventional standard of 75% comprehension i.e., the instructional level. Table 2 also indicates the reading ease (RE) of each passage. All passages were of college grade-level difficulty with passage G ranking most difficult. Both student performance and RE scores indicate a similar rank-order of difficulty for the test passages.

The second research question concerned the proportion of teacher trainees who were functionally literate. According to the NDRT results,

161

Table 1
Mean Cloze Raw Scores on Basal Guide Passages by Groups[a]

	Basal Reading Series				
Group	G	HM	BM	M	Mean
Group 1	11.4	13.8	17.7	17.9	15.2
Group 2	13.5	16.2	18.0	18.8	16.6
Group 3	11.1	12.6	14.2	19.9	14.4
Mean	11.9	14.1	16.8	18.6	15.4

Note. Maximum possible raw score = 41.

[a] Group 1, n=27; Group 2, n=15; Group 3, n=15.

33% of the trainees' reading performance was below the 13th grade-level. Thus, if NDRT performance and RE ratings are utilized, at least 33% of the trainees could be classified functionally illiterate with respect to the sampled teacher guides of basal readers. Use of cloze tests, however, indicates an even greater percentage of functionally illiterate trainees ranging as high as 52%. Determination of the question of literacy rests on the cut-off point utilized in measuring adequate comprehension. If a cloze percentage score of approximately 50% is required for adequate comprehension, then the percentage of functionally literate teacher trainees decreases to 24%.

Table 2
Percentage of Trainees Scoring 40 Percent or Higher on Cloze Passages and Flesch Reading Base Scores of Passages

	Basal Reading Series				
Group	G	HM	BM	M	Mean
Group 1	19	26	74	63	45
Group 2	40	53	67	73	58
Group 3	7	27	54	80	42
Mean	21	34	67	70	52
Flesch RE	39	45	48	47	

The third research question concerned the relationship between professional training (number of reading courses completed) and functional literacy (cloze test performance). It was anticipated that cloze scores and percentages of trainees considered functionally literate would increase as the number of reading courses completed increased. However, the data do not support this supposition. Group 3 subjects, those trainees with the most reading courses completed did not evidence superior test performance as can be noted from inspection of Table 1 and 2.

Discussion

The data indicate that (a) there is a significat difference in the readability levels of sampled teacher guides, (b) despite previous training, the majority of trainees do not adequately comprehend material abstracted from teacher guides, and (c) increasing the number of reading courses completed does not guarantee increased functional literacy levels. Research findings continue to reinforce earlier studies which cite the need for careful evaluation and screening of teachers prior to service. Increasingly it will be necessary for reading experts to look at job tasks and to determine the minimal level of literacy skills necessary for job success. Sticht (1972) found in his research on functional literacy that there was evidence to suggest that easier printed material was more likely to be used and that subjects with relatively low reading accomplishment levels might rely more on asking others or using listening sources rather than reading to accomplish a task.

Training and assessment of future teachers must be relevant and functional, i.e., how to use a teacher guide. Research results point to the need for publishing companies to field test their teacher guides to assure that guides can, in fact, be understood and used as designed. Finally, taking a cue from Sticht's research, probably the material which is easier to read or which is at least within the skill limitations of the potential user will be most used. Therefore, guides should be designed with the skill attainment levels of teachers in mind or teacher education programs must insure that all teacher trainees are functionally literate. Clearly, a case may be established for making functional literacy a requirement for teacher certification.

References

Flesch, Rudolf F. *How to test readability.* New York: Harper, 1951.

Geeslin, Robert H. and York, Patricia W. Literacy skills as a barrier to inservice training. *Journal of Reading Behavior*, 1971, *3*:3, 9–12.

Ranking, Earl F. and Culhane, Joseph W. Comparable cloze and multiple-choice comprehension test scores. *Journal of Reading*, 1969, 193–198.

Sticht, Thomas, G., etal. Project REALISTIC: determination of adult functional literacy skill levels. *Reading Research* Quarterly, 1972, *7*:3, 424–465.

THE NEEDS ASSESSMENT FOR READING TEACHERS: A COMPARISON OF THE DIAGNOSTIC AND PRESCRIPTIVE DIMENSIONS IN READING

MARTHA H. DILLNER
University of Houston–Clear Lake City

CARL N. SHAW
University of Houston–Main Campus

ABSTRACT

All competencies needed to teach reading in the elementary school contain dimensions which involve diagnosing a child and prescribing instruction. Though it would be reasonable to assume that the teacher should first diagnose a child and then prescribe instruction, research at the University of Houston at Clear Lake City indicates that this does not always seem to be the case.

The Needs Assessment for Reading Teachers was developed in order to encompass all the skills thought to be essential to teaching reading into 34 competencies. In this instrument, teachers were asked to rate themselves in the teaching of reading in two ways: their current level on a competency and their desired level on a competency. Each of the 34 competencies were measured on four dimensions: know, use, diagnose, and prescribe for both current and desired levels.

Factor analysis of the responses of approximately 2000 inservice, elementary teachers resulted in four nameable factors present in all four dimensions: 1 = Reading Skills, 2 = Basal Reading Skills, 3 = Organizational and Assessment Skills, and 4 = Peripheral Skills. A scoring key was then generated by which classroom teachers' responses to various dimensions of the teaching of reading competencies could be assessed.

The paper focuses on the diagnose and prescribe dimensions and the implications for the importance of teacher awareness of weaknesses in diagnosis and prescription.

The data analyzed by T scores for the total number of teachers involved illustrates that in general, teachers seem to be diagnosing to the same degree that they are prescribing. The data computed for school x illustrates that these teachers seem to be diagnosing to a much less degree than they are prescribing. The data computed for school y illustrates that these teachers seem to be diagnosing to a much higher degree than they are prescribing. On the whole, the teachers in the University of Houston at Clear Lake City geographical area seem to be

diagnosing and prescribing at about the same level. This would tend to indicate that they are first diagnosing and then prescribing needed instruction.

However, teachers in school x seem to be diagnosing more than they are prescribing. This would tend to indicate that they are over diagnosing. That is, they spend so much time diagnosing that they do not have time to give the child instruction in the reading skills he needs.

Conversely, the teachers in school y, seem to be prescribing more than they are diagnosing. This would tend to indicate that the teachers are not diagnosing as much as they should and therefore, are prescribing instruction on the basis of available materials rather than upon an individual need which they have diagnosed.

A STUDY OF RELATIONSHIPS IN TEACHER PROFICIENCY

WILLIAM R. POWELL
University of Florida

Proficiency in teaching reading may be conceptualized to be a three stage process: acquisition of knowledge, ability to transfer or apply knowledge to varied learning situations (generalizing or problem-solving), and utilization of knowledge in every day teachig tasks (Powell, 1969). The first two stages toward proficiency can be measured by paper and pencil tests designed to tap those functions. The third stage, utilization, has to be assessed through on-the-job observation and rating scales.

The first two stages of this three stage model are essentially the basis of this investigation, namely, what is the relationship between teacher knowledge of reading and teacher application of reading knowledge in simulated classroom reading tasks? Specifically, the questions were: (1) What relationships exist between teacher knowledge of reading and the teacher's ability to solve problems in reading? and (2) What effect does teacher effort have upon teacher knowledge and the teacher's problem-solving ability in reading tasks?

This study was designed to measure and relate three variables: knowledge, problem-solving performance, and teacher effort. The knowledge component was measured by the Artley-Hardin *Inventory of Teacher Knowledge of Reading* (Artley, Hardin, 1971). The problem-solving dimension was assessed through the Powell-Stoll "Problems in the Teaching of Reading" (Stoll, 1971). The Blair "Teacher Effort Scale in Reading" was utilized to determine and estimate of teacher motivation in teaching reading. Each instrument will be briefly described below (Blair, 1975).

The Artley-Hardin *Inventory of Teacher Knowledge of Reading* is a test composed of 95 four-option multiple choice items. The authors state the test has a reliability index of .92. The areas of reading knowledge measured were partitioned into six separate skill areas for purposes of this investigation. They were: instructional goals (15 items), the directed reading activity (13 items), diagnosis and evaluation (22 items), comprehension (13 items), readiness (12 items), and word perception skills (20 items). These categories differ slightly from the content clusters suggested by the test authors. Research by Kingston, Brosier, and Hsu (1975) offers data supporting the validity of the Inventory, and the work of Edelman (1973) contends the results from scores on the Inventory are related to reading gain in children.

The Powell-Stoll test is composed of two forms of nine distinct problem areas: placement levels, grouping, phonics, structural analysis, comprehension, directed reading activity, readability, test data interpretation, and classifying reading errors. The test assumes the presence of information on knowledge and asks the teacher to apply it to specific classroom reading tasks. The test has a total of 86 items and takes about 90 minutes to administer. Stoll (1971) has shown a significant relationship exists between performance scores on this instrument and pupil gain in reading.

The Blair "Teacher Effort Scale in Reading" is a rating device used by an external observer (administrator, supervisor, etc.). It contains four sub-scales entailing effort to: variety of materials, differentiated instruction, teacher-initiated conferences, and record keeping. Blair (1975) found a significant relationship between the degree of effort in the job of teaching reading and children's achievement in reading.

The three instruments previously described were administered to 36 reading resource teachers in a school system in north central Florida. The Artley-Hardin test and the Powell-Stoll test were administered by the investigator of this study. The Blair instrument was sent to each principal or supervisor of the reading teacher included in this study for his or her completion and evaluation. While each of the teacher subjects was designated as a reading resource teacher in her respective school, there was a wide range of training and experience in the group. Experience ranged from one year to many years in teaching, and the level of training varied from just over a bachelor's degree to a few with several graduate hours in reading past the master's degree.

The data from each test and its sub-tests were correlated with all other variables in the study. There were 22 variables: seven from the Artley-Hardin Inventory, 10 from the Powell-Stoll problems test, and five from the effort scale. The intercorrelation matrix is shown in Table 1. Any coefficient of .33 or greater represents the .05 level of confidence, while a figure of .42 or greater is significant at the .01 level.

An inspection of the data reveals a significant relationship between the total knowledge score in reading and the ability to solve classroom reading problems. However, the relationships between these two variables was .63; and while it is clearly significant well beyond the .01 level of confidence, the degree of strength between the two variables would only be considered to be moderate. A further analysis of the matrix will reveal that many of the sub-tests' correlations are low and/or not significant. For example, both the Artley-Hardin and the Powell-Stoll tests purport to measure the areas of the directed reading activity, comprehension, and word perception skills (phonics and structural analysis). Yet the relationship between knowledge and application of knowledge in these three skill areas was only .43, .28, and .54 and .28, respectively. One would have normally hypothesized a

Table 1. Intercorrelations of Teacher Knowledge, Problem-Solving, and Effort Measures

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1.00	.65	.55	.35	.38	.52	.75	.16	.09	.54	.16	.05	.43	.13	.22	.30	.41	.07	.28	.40	.16	.19
2		1.00	.61	.29	.31	.48	.73	.05	.03	.33	.23	.11	.39	.06	.12	.24	.32	.23	.28	.45	-.06	.27
3			1.00	.57	.38	.53	.79	.08	-.06	.47	.37	.24	.43	.26	.42	.40	.54	.05	.11	.47	-.20	.13
4				1.00	.62	.67	.77	.30	.12	.47	.47	.28	.50	.23	.29	.48	.61	.02	.09	.29	-.07	.09
5					1.00	.50	.68	.03	-.03	.27	.19	-.03	.26	.09	.11	.14	.21	.16	.01	.01	-.19	-.10
6						1.00	.82	.15	.35	.54	.28	.27	.50	.08	.40	.61	.67	.30	.39	.37	.07	.33
7							1.00	.20	.13	.56	.36	.24	.53	.15	.36	.50	.63	.13	.26	.45	-.10	.22
8								1.00	.07	.03	.16	.19	.12	-.01	.12	.20	.32	.33	.47	.33	.12	.37
9									1.00	.47	.11	.18	.23	-.04	.17	.29	.50	.01	.14	-.13	.02	.01
10										1.00	.41	.09	.47	-.04	.34	.37	.66	.04	.24	.40	-.01	.19
11											1.00	-.04	.42	-.06	.50	.17	.58	.03	.02	.32	-.12	.06
12												1.00	.19	.30	.23	.30	.45	.07	.10	.15	.13	.14
13													1.00	-.11	.16	.23	.53	.07	.26	.21	-.24	.04
14														1.00	.34	.09	.25	.01	.09	.27	.16	.17
15															1.00	.65	.77	.16	.22	.32	-.09	.18
16																1.00	.74	.29	.26	.27	-.14	.19
17																	1.00	.15	.34	.43	-.06	.26
18																		1.00	.73	.55	.57	.86
19																			1.00	.60	.40	.82
20																				1.00	.43	.80
21																					1.00	.76
22																						1.00

higher degree of correspondence between knowledge into simulated classroom problems. However, an examination of the type and number of items in the two instruments might likely explain these levels of relationship.

The relationship between knowledge and effort was only .22, and between problem-solving ability and effort, only .26. Both correlations are very low and insignificant. A cautionary note must be extended here regarding the results of this instrument as used in this study. The investigator has the distinct impression that the scores on the Teacher Effort Scale as collected for this study should be held suspect. The mean scores on the sample in this investigation were significantly higher than the scores obtained by Blair (1975) in his validation study. The investigator suspects that the principal or the supervisor rating the teachers on the effort scale did not know the depth of the activities and scored the teachers using a global impression of them as a teacher, rather than a specific observed knowledge of their teaching proficiency. Too many of the teachers were given unusually high ratings on the twelve items on the scale.

A factor analysis was performed on the 22 variables used in this study to determine the degree to which the instruments measured the separateness of components as suggested by the structure of their content. Eleven factors were found, and these eleven components accounted for 95 percent of the total score variance. The factors are shown in Table 2.

It is apparent from Table 2 that two factors emerged from the Artley-Hardin *Inventory of Teacher Knowledge of Reading*: one cluster consisting of the tests presumably measuring the areas of instructional goals, the directed reading activity, and diagnosis and evaluation content; the second cluster grouped together the areas of comprehension, readiness, and word perception skills. The first cluster might be labeled as a knowledge of the process characteristics of teaching reading, and the other factor called a knowledge of reading content (or product factors). However, this finding is in contrast to the study by Kingston, Brosier, and Hsu (1975), who found no identifiable clusters in the Inventory.

The data presented here would indicate that the Artley-Hardin *Inventory of Teacher Knowledge of Reading* is a single or two-factor instrument which measures selected areas of reading knowledge. While the test may differentiate between levels of obtained knowledge in the reading area, it offers no diagnostic and prescriptive qualities, nor much information for self-evaluation and direction for improvement. Further, if application of knowledge is a criterion of performance, the scores from the Artley-Hardin offer little in the way of predictive power as to who will apply the knowledge they possess.

Table 2. Sub-test Loadings from Factor Analysis of Teacher Knowledge, Problem-Solving, and Effort Measures

1	2	3	4
Instr. Goals (A-H) DRA (A - H) Diag. & Eval. (A - H)	Var. Materials (B) Diff. Instr. (B) Tchr. Conf. (B) Rec. Keeping (B)	Groups (P - S)	Readability (P - S)
5	**6**	**7**	**8**
Structural Anal. (P-S)	Placement (P-S)	Test Interpret. (P-S) Class. Errors (P-S)	Comprehension (A-H) Readiness (A-H) Word Perception (A-H)
9	**10**	**11**	
Comprehension (P-S)	DRA (P-S)	Phonics (P-S)	

The Powell-Stoll instrument does measure teacher proficiency in solving selected problems as might be encountered in a teaching situation. It shows a moderate relationship to teacher knowledge and has separate components which offer possible potential for diagnosis and prescription for teacher education programs.

The data in this study would suggest that the three stage model of studying teacher proficiency has merit. It is possible to obtain measures of teacher performance in the knowledge domain, problem-solving or application areas, and estimates of teacher effort in a given performance area such as reading.

There is clear evidence that there is a significant and marked relationship between the teacher's knowledge in a given performance area and his/her ability to solve problems in that area. Support is given for the old adage that "one cannot teach (or apply) what one does not know." However, the moderate nature of this relationship would suggest an evaluation of information or knowledge taught and learned as it may have little or no applicability. Another age-old question persists: "What knowledge is worth knowing?"

The relationship of knowledge and problem-solving ability with teacher effort in a given area deserves further exploration. The evidence in this study indicates little relationship between these abilities. Yet, the utilization of knowledge would seem to demand that effort be extended to realize the potential of knowledge and its application.

Teacher education programs in reading will need to refine techniques for the diagnosis and prescription of teacher proficiency. If instruments can be devised, developed, and refined to assess each stage of progress toward proficiency, then programs can be designed to meet the needs of teachers in various stages of development.

REFERENCES

Artley, A. S. and Hardin, V. B. *Inventory of Teacher Knowledge of Reading.* Columbia, Missouri: Lucas Brothers Publishers, 1971. (A 1975 edition is now available.)

Blair, T. R. Relationship of Teacher Effort and Student Achievement in Reading. Unpublished doctoral dissertation, University of Illinois, 1975.

Edelman, E. Levels of Teacher Reading-Knowledge and Pupil Initial-Status Reading Achievement: Their Relation to Levels of Pupil Residual-Reading Achievement Gain. Unpublished doctoral dissertation, University of Missouri–Columbia, 1973.

Kingston, A. J., Brosier, G. F., and Hsu, Y. The Inventory of Teacher Knowledge of Reading–A Validation. *The Reading Teacher*, 1975, *29*, 133-136.

Powell, W. R. The Effective Reading Teacher. *The Reading Teacher*, 1969, *25*, 603-607.

Stoll, P. D. The Construction and Initial Validation of An Instrument Designed to Assess Problem Solving Proficiency in the Teaching of Reading. Unpublished doctoral dissertation, University of Illinois, 1971.

Turner, R. L. and Fattu, N. A. Skill in Teaching, A Reappraisal of the Concepts and Strategies in Teacher Effectiveness Research. *Bulletin of the School of Education,* Indian Effectiveness Research. *Bulletin of the School of Education,* Indiana University, 1960.

Tests and Index Numbers for Table 1

Artley-Hardin Inventory of Teacher Knowledge of Reading (A–H)

1. Instructional Goals
2. Directed Reading Activity (DRA)
3. Diagnosis and Evaluation
4. Comprehension
5. Readiness
6. Word Perception Skills
7. Total Knowledge Score (Sum of 1–6)

Powell-Stoll Problems in Teaching Reading (P–S)

8. Placement
9. Grouping
10. Phonics
11. Structural Analysis
12. Comprehension
13. Directed Reading Activity (DRA)
14. Readability
15. Test Data Interpretation
16. Classifying Reading Errors
17. Total Problem-Solving Score (Sum of 8–16)

Blair Teacher Effort Scale in Reading (B)

18. Variety of Materials
19. Differentiating Instruction
20. Teacher Initiated Conferences
21. Record Keeping
22. Total Effort Score (Sum of 18–21)

HOW NECESSARY ARE CERTAIN SKILLS FOR READING COMPETENCE?

MARIE CARDUCCI-BOLCHAZY
State University of New York at Albany

ABSTRACT

Reading management systems are based on the concept of prerequisite skills. For this study specific skills from the word attack component of the *Wisconsin Design for Reading Skill Development* were selected to invistigate whether these skills are indeed prerequisite for competence in reading. If there are competent readers who lack these skills, it may be concluded that the skills are not prerequisite to reading competence.

The literature generally is supportive of the teaching of consonants and their variant sounds and three letter consonant blends. The literature contradicts the view that skills in applying the following rules are prerequisite to reading competence: (a) two vowels together, (b) final vowel, (c) rules for dividing between syllables. There is a little evidence supporting the teaching of accenting, but this evidence is far too scant to suggest that the skill to recognize and place accent marks is essential for reading competence. In regard to the use of nonsense words on tests of word attack skills, there is evidence that some children who can pronounce real words still have difficulty pronouncing nonsense words. McNeill's (1974) study on false prerequisites was also reviewed. McNeill found the following skills to be false prerequisites: (a) selecting words with affixes, (b) distinguishing meanings of homographs, (c) selecting similar sounds or r-controlled vowels, and possibly, (d) selecting pairs of consonant variables.

For the first part of this study, 25 eighth-grade students who had scored at the 50th percentile or higher on a standardized reading achievement test were tested on five subtests of the *Wisconsin Design for Reading Skill Development–Word Attack*, levels C and D. These students attended an inner-city school located in an area of low socioeconomic level. The racial composition of the sample was mixed. The experiment with the eighth-graders did not yield any evidence against the teaching of three letter consonant blends, the two vowels together generalization, or syllabication. However, there were data to contradict the view that skills in applying the final vowel generalization and in selecting correctly-accented words are prerequisite to reading competence.

For the second part of the study, this investigator surveyed 21 third-, fourth-, and fifth-grade teachers who had been using the word attack component of the *Wisconsin Design*. The teachers were asked to estimate the percentage of their competent readers who could pass particular subtests on the *Wisconsin Design*. The data from this survey support the teaching of consonants and their variant sounds and three letter consonant blends. The data from this survey also provide some evidence to contradict the view that skills in the following areas are prerequisite to reading competence: (a) final vowel generalization, (b) long vowel sounds, (c) two vowels together generalization, (d) silent letters, (e) syllabication, and (f) accenting.

This study provides some evidence that not all skills taught in reading management systems are prerequisite to reading competence.

READING ATTITUDE AND ACHIEVEMENT OF TEACHER EDUCATION STUDENTS

LAWRENCE ERICKSON
West Virginia University

THOMAS HATCHER
West Virginia University

ABSTRACT

The McGraw-Hill Basic Skills System Reading Test and 70 Item Attitude Instrument were administered to 171 teacher education students at West Virginia University. Results indicated few significant differences in reading attitudes and achievement by sex, major, or level of training although males and graduate students read faster than females and undergraduate students. An analysis of percentages of students scoring by stanines led to the recommendation that teacher training programs should contain components which attempt to involve all prospective teachers in activities designed to improve their own reading abilities. This seems especially true for reading specialists since their primary task is often improving the reading skills of others. The point is that teachers' attitudes and abilities regardig improving reading skills in others would seem to be directly related to the experiences they have at improving their own reading abilities.

For example, during a reading methods class, prospective teachers could take a reading test, set individual goals, and work to improve their own reading skills with the guidance of an instructor who is familiar with adult reading skill improvement.

Part of a teacher training program should include the administration of a reading test to undergraduate education majors. Minimum reading standards should be established and prospective teachers who fall below these standards should be referred to a reading improvement laboratory to insure that reading instruction is carried out by teachers who demonstrate above average reading abilities.

Additional research is needed to clarify the relationship between reading proficiency and teaching proficiency. Also, there is a need to establish more valid scales for measuring the reading attitudes of teachers. And finally, longitudinal studies which assess the reading abilities of prospective reading specialists and follow them through graduate programs which include a reading improvement component would shed some light on the whole question of the reading abilities of reading teachers.

SYMPOSIUM: READING PROBLEMS AND HABITS OF AMERICAN INDIAN YOUTH

EMERY P. BLIESMER
Pennsylvania State University

INTRODUCTION

With the current and long-needed stress on and efforts to help minority or given ethnic groups attain consideration and equality comparable to that given other groups in our country, a symposium session on reading problems and habits of American Indian youth seemed quite timely and appropriate for this year's conference. It is also very fitting that the major presentations for this session have been prepared by members of native American Indian tribes.

Consideration of the topic from the lower or elementary grades level has been done by Sandra J. Fox, a Sioux Indian, who is currently an Education Specialist with the Aberdeen, South Dakota, Area Office of the Bureau of Indian Affairs (BIA). Her main recent work has involved evaluation of Title I ESEA programs in the BIA schools in the Aberdeen Area, which involves parts of both North and South Dakota. Before her current position, Mrs. Fox served for a time as a BIA language arts consultant in the Albuquerque, New Mexico, Area; and prior to that she was an English teacher in the Eagle Butte, South Dakota, schools. (It might also be added that Mrs. Fox's mother is a teacher in Indian schools in Alaska.)

Treatment of the topic from principally the secondary student level has been done by Billie Sue Allen, a member of the Creek Indian tribe, who is presently a reading and English teacher in the Riverside Indian School, Anadarko, Oklahoma. Before she held this position, Mrs. Allen was an English and reading teacher in the Fort Sill (Oklahoma) Indian School. Both Mrs. Allen and Mrs. Fox have been pursuing doctoral programs at Pennsylvania State University, in the major field of Academic Curriculum and Instruction. Both have completed the course work of their programs, and are presently working on their doctoral dissertations.

Treatment of the topic from the viewpoint of the college student or young adult level has been given consideration by John W. Tippeconic, a member of the Comanche/Cherokee tribes, who has been a social studies and mathematics teacher in both BIA and public junior and senior high schools in Arizona and New Mexico. He is now vice president of the recently built and opened Navajo Community College in Tsaile, Arizona. Mr. Tippeconic is one of the young men and women selected to participate in the Native American Leadership Program at Pennsylvania State University, an ongoing program of the past several years which has been sponsored jointly by the university and the Bureau of Indian Affairs. (Mrs. Allen's and Mrs. Fox's husbands are also participants in this program.) During this academic year, Mr. Tippeconic is on leave of absence from his Navajo Community College position while he completes work on his doctoral dissertation and also serves as associate director of the Native American Leadership Project.

SECONDARY INDIAN YOUTH AND THEIR READING

BILLIE SUE ALLEN
Riverside Indian School

Gray (1963) describes reading as a four-step process: perceiving the printed word, comprehending its meaning, reacting to the concept involved, and integrating what it says into one's total experience. Each of these steps reaches back into the meaning background of the individual, hence, his or her culture. Roessel (1962) asserts that the Indian children often find reading hard because their language ability is limited and because their experiential background is distinctly different from that of the non-Indian American children for whom most reading situations are planned.

Bryde (1967) defines culture as a common way of life of people. Different people have different ways of life and come up with different answers to the same problems facing them. The different answers, or how a group of people reacts to these problems, are their "culture." Parents teach their children ways of doing important things so that, when their children grow up, they will in turn teach their children the same ways of doing the same important things. In this way, a culture is handed down from generation to generation.

Perhaps the most important fact to bear in mind when examining the cultural background of the American Indian is that Indians are not immigrants to this country. The root and home of their culture is here, and their resistance to assimilation into the larger American culture cannot be compared with that of others who have left the mainstream of their own cultures and freely chosen to live in a different one.

Townsend (1963) studied the reading achievement of eleventh- and twelfth-grade Indians in New Mexico. He found for this group that "Indian students generally achieve at least five years below grade level." Study results also indicated that public schools fail to function equally well for all students in the area of reading and that teachers and principals may not be aware of the many causes of retardation in reading among ethnic groups.

Philion and Galloway (1969) observed that Indian children do not achieve as well as non-Indian children in the areas of vocabulary, comprehesion, and speed of reading. As time progressed, Indian children fell further and further behind non-Indian children in reading achievement. The study indicated that reading programs designed essentially for non-Indian children were not adequate for Indians.

Byrde (1970) hypothesized that conflict between the white and

Indian cultures comes to focus at adolescence and causes severe personality disturbances which block achievement. Having identified the central pattern – alienation and anomie, with the resultant feelings of rejection, depression, and anxiety – it was seen that Indian youth are alienated from themselves and others. They are not effectively identified with their Indian heritage, nor can they identify with the hostile, white world facing them. They are, during the troubled years of adolescence, "nothings." They have extremely crippling negative self-images. They have no direction to their lives and they are lost.

Roessel (1962) writes that the Indian today lives between two worlds. One is the traditional way of life represented by the culture of Indians prior to the coming of the white man. The other is the modern American civilization with its own distinctive patterns and sets of values. In order for educators to reach the Indian child and to instill within that child a desire to learn, it is necessary for educators to have a basic understanding of the culture of the Indian.

Some Problems Displayed by Indian Students in Reading

1. Irregular attendance patterns with resulting disruptions and lack of continuity in their work. Failing students become disheartened and develop resentments toward school, toward adults, and toward the "Anglo way" of living. They become emotionally distrubed and search for satisfactions elsewhere.
2. Feelings of insecurity and inadequacy possibly growing out of educational, economic, and social disadvantage as compared to their non-Indian peers.
3. Reticence and shyness resulting in avoidance of class participation whenever possible. Problems in enunciating the English language would be one cause of reticence.
4. Reading test scores have indicated that Indian students do not achieve satisfactorily in reading, especially in vocabulary and comprehension, possibly because of very limited and narrow concept development, or limited environmental experiences.
5. Language deprivation is very apparent. Indian students are deprived of that part of language that can only be acquired through verbal teaching–the knowledge, the meaings, the explanations, the ability to question in search of information. The students have spent their childhood in an environment in which verbal teaching did not frequently take place and where language with which verbal teaching is carried out was not used.

Some Points One Might Remember in Working with Indian Students

Each teacher must remember the old education motto, "Who would dare to teaching must never cease to learn." Those who want to avoid frustration in teaching Indian pupils had better begin to learn.

Martin and Castaneda (1970) suggest that in working with bilingual students one needs to remember that no greater compliment can be paid to a child, no greater indication of his or her worth can be given, than that a teacher values the child's thoughts highly enough to wait and to attend to the child while he or she puts them into words; and no greater encouragement or motivation can be offered to the child than the knowledge that he or she has communicated.

York and Ebert (1970) believe that a teacher should translate the values of the home in such a way that the child continues to respect and appreciate his or her home and family while learning the values of the school and of the majority of society. By teaching alternative, acceptable, and appropriate forms rather than being critical of those used in the home, the teacher is helping the child to live in two worlds–the home and the school.

A good way to establish rapport with Indian students is to become acquainted with their means of subsistence, by traveling and visiting on the Indian reservations, by talking to someone who knows about Indian customs and beliefs and including items that relate to the tribes in school work and class projects, and by becoming familiar with typical Indian food and being able to say that one has eaten some.

Roessel (1962) points out that the extended family structure of the Indian gives the children a strong feeling of security. The teacher should know something about the clan system of the tribes, know something about the organization of the family of the tribes, respect the Indian method of determining kinship, and never ridicule.

The teacher must have some understanding of the role of Indian religion in the lives of those he or she is teaching. The teacher should become familiar with local Indian dances and ceremonies, respect the Indian student's right to be absent from school to attend important ceremonies, be careful not to knowingly or unknowingly cause a child to lose face in regard to his or her traditional religion, and invite the medicine men to the school and have them address the students in regard to myths and legends. These myths and legends might be used in reading and storytelling.

The effective Indian educator is one whose concern for these children extends beyond classroom activities to citizen participation in efforts to upgrade the neighborhood, to abolish discriminatory practices, and to support social action for human betterment. Such a teacher will have an impact on his or her pupils and will engage them to the end that they realize the talents they have in the same measures as children elsewhere–talents which they cannot develop without the teacher's help. Eisenberg (1967) states that the real challenge is to lead forth the children to a place in the sun, because they are not learning by themselves, and they are not learning in the way society has been teaching them. If teachers accept this challenge, use their ingenuity, and have faith in children's capacity for growth, teachers will succeed, and society will owe them a debt that money and prestige can never repay.

References

Bryde, J. F. *Acculturational Psychology or Modern Indian Psychology.* Washington, D.C: U.S. Department of the Interior, 1967.

Bryde, J. F. *A Study of Scholastic Failure and Personality Conflict.* Vermillion, South Dakota: South Dakota Press, 1970.

Eisenberg, L. Strength of the Inner City Child. In Passow, A H. Goldberg, M. and Tannenbaum A. J. (Eds.) *Educating the Disadvantaged* New York: Holt, Rinehart and Winston, Inc., 1967. Chapter 5.

Gray, L. *Teaching Children to Read.* New York: The Ronald Press Company, 1963.

Martin, C. I., & Castaneda, A. M. Nursing School and Kindergarten. In Horn, T. D. (Ed.), *Reading for the Disadvantaged.* New York: Harcourt, Brace and World, Inc., 1970. Chapter 8.

Phillion, W. and Galloway, C. Indian Children and Reading Programs, *Journal of Reading,* 1969, *2,* 553-560.

Roessel, R. A., Jr. *Handbook for Indian Education.* Los Angeles: Amerindian Publishing Company, 1962.

Townsend, I. D. Reading Achievement of Eleventh and Twelfth Grade Indian Students, *Journal of American Indian Education*, 3 (October, 1963), 9–10.

York, L. J., & Ebert, D., Primary Level: Grades 1–3. In Horn T. D. (Ed.), *Reading for the Disadvantaged.* New York: Harcourt, Brace and World, Inc., 1970. Chapter 9.

IMPLICATIONS FOR THE TEACHING OF READING TO AMERICAN INDIAN STUDENTS IN THE ELEMENTARY GRADES

SANDRA J. FOX
Bureau of Indian Affairs

Studies such as the National Study of American Indian Education (Havighurst, 1970) and the Senate Subcommittee Report on Indian Education (U.S. Congress, 1969) indicate that American Indian students generally score less well on tests of school achievement than do white students. Further, these studies report that Indian children score especially low in the area of reading. Teachers of Indian students must attempt to remedy this situation.

The teacher of Indian children in the elementary grades is faced with special challenges. Many American Indian children come to school with limited knowledge of the English language. The Senate subcommittee report states that one-half to two-thirds of all Indian children enter school with little or no skill in the English language. Indian children often come from homes in which the tribal culture is still very strong and is quite different from the non-Indian culture of the school setting.

It is not surprising that the literature contains, repeatedly, the need to utilize reading materials and methods which reflect the language and culture of Indian children, especially for beginning reading instruction. This is a wonderful ideology, but it is not so easy to implement. The reasons for this are many.

Reading Materials

There are very few materials on the market designed especially for use in teaching reading to Indian children. Because of the wide diversity of cultures, of degrees of acculturation, of Indian languages, and of knowledge of English among Indian people, it is not economically feasible for commercial materials developers to produce materials for Indian groups. Neither is it feasible to promote one reading program, method, or set of reading materials for Indian students.

There are at least 200 different groups of Indian people in the United States. It could be said that each tribe should have reading materials designed for use in its schools, but within each tribe there are differences among members in regard to language usage and culture. Each classroom, therefore, requires varied reading materials based upon the cultures, the degrees of acculturation, and the language abilities of its students. This places a tremendous burden on teachers, who must do a great deal of research, gathering, planning, and preparing of materials to suit the students' needs if they really want to delve into solving the problem of reading deficiencies among Indian students.

Teachers

Most teachers of Indian students are non-Indian. They usually are not familiar with the Indian culture or language of the students they teach. Teachers of Indian students are generally not trained well enough to deal with Indian languages in the classroom or even to detect aspects of the local Indian dialect of English, much less use them to educational advantage.

When Indian students come to school speaking English, teachers usually mistakenly believe that there are no language differences to take into account. They fail to realize that there are English idiom and vocabulary elements which may be entirely foreign to Indian students. Likewise, teachers fail to recognize Indian slang and other aspects of "Indian English" used in the Indian community. If a teacher does recognize the varying aspects of the English spoken by Indian students, he may approach them on the basis of the "deficit" theory rather than the "different" theory. Teachers also fail to acknowledge Indian cultural aspects and tend to assume that Indian students want to be or should be acculturated or that Indian students are more acculturated than they may actually be.

School Systems

School systems for Indian students, except in a few cases, are no different than other school systems in the United States in regard to curriculum and course content. The money spent and the effort extended are for the purpose of providing services which meet the state requirements and/or provide for some special kind of accreditation. The feeling is, generally, that there is not enough money to provide for the inclusion of Indian curriculum in the schools. Where there is an Indian curriculum, it is usually provided through federal title program funds. Most school personnel probably feel that they have neither the time nor the money to devote to providing the kind of reading program which seems necessary for Indian children.

An important element that is lacking in most Indian school systems is parental involvement. Only recently have Indian parents been given the opportunity to participate in the education of their children in Bureau of Indian Affairs schools. Indian parents, then, are very hesitant to participate in education activities and most often look upon the school and its programs as foreign and threatening. The involvement of Indian parents in the education of their children is seen as one of the most necessary steps to be taken toward increasing the children's achievement in school. The fact that Indian parents may not take an active role in their children's reading experiences in the elementary grades is an important aspect to consider and to work toward rectifying.

ESEA Title I

A federal program, Title I of the Elementary and Secondary Education Act (ESEA), provides funds for remedial programs to be designed to increase the reading skills of children. There are Title I programs in all Bureau of Indian Affairs schools and in some public schools where Indian children are in attendance. In the early elementary grades, these Title I programs are often designed to prevent Indian students from falling behind in reading. Unfortunately, the same problems exist with Title I programs as with regular school reading programs for Indian students. Thousands of dollars are spent on commercially designed diagnostic-remedial materials and machines and reading specialists are hired, but the materials are usually those written for non-Indians and the reading specialists don't know any more about the children they are to teach than do the regular teachers. Title I has not been the answer, then, for solving the reading problems among Indian students.

Worthwhile Recommendations

Some of the recommendations concerning the teaching of reading to Indian children in the elementary school are well worth considering.

Robinett (1970), known for his work on the *Miami Linguistic Readers*, stresses that there is a wide diversity among American Indians in their ability to use the English language. He feels that teachers need to use approaches which are best suited for their students. He recommends that teachers should draw from several approaches in order to obtain a comprehensive and controlled program. Some of the approaches from which a program could be formulated are the language-experience approach, the second language approach, and the sound-symbol correspondence approach. Fearn (1971) recommends the use of the language-experience approach for Indian students and reports on the results on the Navajo reservation. He reports that at Lukachukai, on the Navajo reservation, the language and environment of the children are capitalized upon for reading instruction and that desirable results have been obtained.

Zintz (1971) recommends that, whenever possible, Indian students in the bilingual approach, should learn to read first in their native languages. There are now a number of bilingual education programs for Indian students, mainly in the Southwest; however, these programs have not progressed to the point that Indian students have learned to read first in their native languages and then in English, primarily because materials have had to be developed.

Evaard and Mitchell (1966) have analyzed some of the beliefs and values encountered in the Scott-Foresman Basic Readers and have found that they conflict with Indian beliefs and values. They strongly urge that the materials used to teach Indian children to read should reflect the culture of the Indian child rather than that of the middle-class urban child.

Among Griese's (1971) recommendations concerning reading comprehension is the suggestion that reading materials which deal with the cultural setting of the Indian child should be used. When the child can relate emotionally to the story situation, his imagination can work in abstractions and may become involved in complex cause and effect thinking. The teacher will have to guide the students in this thinking process. Griese also suggests teacher-reading to the class as an important activity.

Narang (1974) stresses, among other things, the importance of teacher understanding and acceptance of Indian children and the need to provide success experiences for them, in reading. Rich (1973) includes the need to recognize individual differences among Indian children in regard to reading abilities and to take into account cultural differences and individual characteristics of students.

Conclusion

The problem of reading deficiencies among Indian students cannot be ignored. The National Advisory Council on Indian Education (1974) has recommended that the Bureau of Indian Affairs and all other federal agencies dealing with Indian education be required to publicly declare a goal of closing the education gap between Indians and other Americans by 1976.

The teaching of reading to Indian children in the elementary school requires the inclusion of all of the facets of any good reading program plus the inclusion of emphasis on the Indianness of the children. In order to provide good reading programs for Indian children, school systems will have to shift priorities and teachers will have to work extra hard to provide the best possible reading instruction.

References

Evvard, E., & Mitchell, G.C. Sally, Dick and Jane at Lukachukai. *Journal of American Indian Education*, 1966, *5*, 2-6.

Fearn, L. Report of three pilot studies into initial reading: Navajo children. *Elementary English*, 1971, *48*, 390-394.

Griese, A. Focusing on students of different cultural backgrounds: The Eskimo and Indian pupil – Special problems in reading comprehension. *Elementary English*, 1971, *48*, 229-234.

Havighurst, R.J. The education of Indian children and youth. *National study of American Indian education* (Series 5, No. 6). Chicago: University of Chicago, 1970.

Narang, H.L. Improving reading ability of Indian children. *Elementary English*, 1974, *51*, 190-192.

National Advisory Council on Indian Education. *First annual report to the Congress of the United States from the National Advisory Council on Indian Education.* Washington, D.C.: Government Printing Office, 1974.

Rich, G. Teaching reading to the American Indian. Paper presented at the annual meeting of the International Reading Association, Denver, 1973.

Robinett, R.F. The teaching of English reading to American Indian children. *English for American Indians*, Spring 1970, 1-11.

U.S. Congress, Senate, Special Subcommittee on Indian Education. *Indian education: A national tragedy – A national challenge.* Washington, D.C.: Government Printing Office, 1969.

Zintz, M.V. What classroom teachers should know about bilingual education. In *Bilingual education for American Indians*, Washington, D.C.: Bureau of Indian Affairs, 1971.

OBSERVATIONS ON THE AMERICAN INDIAN COLLEGE STUDENT, WITH IMPLICATIONS FOR READING

JOHN W. TIPPECONNIC
Navajo Community College

Background

American Indians are attending colleges and universities in increasing numbers. Approximately 20,000 American Indians are currently pursuing degrees in institutions of higher learning. This represents an increase of approximately 17,000 students since 1957. However, the increase does not necessarily result in a higher number of graduates. In 1969 it was reported that only 3 percent of all American Indian students who enter college graduate. The national average is 32 percent. One in every 100 Indian college graduates will receive a master's degree (U.S. Congress, 1969). The problem remains, despite the fact that these figures were compiled over five years ago.

Why the high drop-out rate? It is a very complex problem with a multiplicity of factors. The drop-out rate has been related to teacher expectation, inadequate counseling, inadequate academic preparation, study habits, motivation, emotional and social adjustment problems, and insufficient funds. English language difficulties have also been identified as a major factor. Research indicates that English skills are definitely correlated with success in college (McGrath, 1962). Spencer (1973) states that the average drop-out is below grade average in reading ability and that the reading problem is the main reason that students quit school.

To direct a program toward the reading problem alone would be unfruitful, just as it would be to attack any one of the other variables related to the drop-out problem. The factors are interrelated and cannot be effectively treated separately. A global approach must be used, realizing the total environment of the student.

The American Indian stereotype promoted by movies, television, and textbooks is a myth. Nothing could be further from the truth. The American Indian is a very diverse group. Today there are over 800,000 American Indians, Aleuts, and Eskimos representing over 300 major tribes, living in all 50 states. Approximately 450,000 Indians live on reservations; the remaining 350,000 live off-reservation, many in urban areas. Over 300 different languages are spoken, and each tribe has its own way of life, religious beliefs, customs, and geographic region. Tribes vary in their degree of assimiliation into the dominant society. Differences exist in the amount of awareness and participation in the non-Indian world because of geographic and social isolation. These differences must be recognized and accepted before educational programs, including language programs, can be successfully implemented.

The American Indian College Student

The diversity among tribes also applies to students. There is no typical American Indian college student. Not all students have problems unique to Indians. Many have no problems at all and possess the ability to adjust socially and compete academically in any institution of higher learning. Indian students can be found in all types of colleges and universities located in all areas of the country. The community colleges, with their open door policy and ability to provide developmental programs, are enrolling more Indian students yearly.

A closer look at some of the envoronmental considerations will give a better understanding of the college student. Reading behaviors and habits of the college student are also implied. Students come from two geographic locations, reservattions and off-reservation areas. Reservations have helped to confine and preserve tribal languages and customs. As a result, many students from reservations, especially in the Southwest, grow up in geographic and social isolation, often with limited experience with the non-Indian society. By sheer exposure, students from off-reservation areas are more familiar with the English language. Interaction with non-Indians enhances reading by providing opportunities to improve the communication skills of listening, speaking, and writing.

It is a known fact that socioeconomic conditions are related to the achievement of students. Poverty and illiteracy of many Indian families result in a lack of exposure to the reading-related skills of speaking, listening, and writing the English language. The accessibility of books, magazines, and newspapers is limited, and such materials are often found only outside the home. The Navajo Tribe (1973) reports that 60% of the homes on the reservation lack electricity, denying necessary light required for reading activities.

Reading in the schools is often limited to academic class activities. Very little reading is done for pure enjoyment. If a student does have an interest in reading, he is often discouraged by peer group pressure. He is labeled and may become ostracized by his peers. This is especially true at the elementary and secondary levels and gives an indication as to how the reading habits and behaviors of the college student are formed.

English is a second language for many students. Nearly a third of the 120,000 members of the Navajo Tribe are functional illiterates in English (U.S. Congress, 1969). Children enter school speaking only their native language; by the time they leave high school, the majority are below average in reading ability. The U.S. Commission on Civil

Rights (1972) reports that by the twelfth grade, Indian students are 2.7 years behind Anglo students in reading ability. In New Mexico, Indian students exhibited the lowest achievement levels in reading. Again, it should be emphasized that this is not common to all Indian students–some Indian students speak no native language and others are somewhat fluent in both English and their native language.

One of the most frequently mentioned concerns facing the American Indian college student today is the problem of adjusting to the college environment. Artichoker and Palmer (1959) contend that study habits, academic background, limited vocabulary, and slowness in reading rate are definite factors. They further state that these characteristics are more prevalent in bilinqual students.

A possible explanation for the adjustment problem is the conflict between the values, attitudes, and life style of the Indian and the non-Indian college environment. Supportive evidence is provided by Jensen and Jensen (1970) when they state that the fears of the new environment were great enough to prevent Indian freshmen from remaining on campus.

Numerous efforts have been made to reduce the high drop-out rate. Upward Bound programs, emphasizing skills in reading and writing and the development of the thought process have been implemented. Short term concentrated programs in English language, guidance, and counseling have been conducted. Ethnic and Indian studies programs have been established at many colleges and universities to assist students. No doubt, these efforts are noteworthy and will result in more graduates and a reducation of the drop-out rate.

A Global Approach

However, the creation of a learning environment based on the background of the student has important implications for improved reading activities and reduced drop-out rates. The home environment and adjustment and language considerations, which were previously mentioned, are viewed as building blocks and not barriers to educational success. The recent and developing Indian controlled schools have an opportunity to create this kind of environment.

Within this movement is the Indian controlled and operated community college system. Inherent in the community college concept is the opportunity to provide a comprehensive curriculum, an open door admission policy, and a community–oriented college in all aspects and practices.

The key is the community-oriented aspect. The Indian community college can create an environment conducive to a bilinqual-bicultural program that recognizes and accepts the Indian culture as equal and coexistent with the non-Indian culture. Opportunities for formal and informal experiences, promoting Indian heritage and culture, are provided by members of the community and institution. The community becomes the classroom. If one chooses, the Indian influence can be felt in every facet of the institution. Why not present a math concept in a native language? Why not have the human anatomy instructor teach the parts of the body in a native language as well as English? Why not serve Indian delicacies in the cafeteria? The result is that a student becomes familiar with his own culture. Concurrently, he learns the necessary skills to function in the dominant society. The development of pride in one's culture occurs. Self-awareness leads to an improved self-concept which in turn influences achievement. Bass (1971) found that self-concept of Indian students was significantly correlated to academic achievement, including reading.

What implications does this have for reading? Reading is not a group activity. Reading instruction is individualized in such a program, individualized in the sense that the environment is relevant to the reader and becomes functional in the reader's own terms. It is conceivable that motivation and interest develop because the reader becomes directly involved in the daily concerns of his or her people. The thought process develops, formulating ideas that are shared through the communicative skills of speaking and writing.

Listed below are a number of practical activities in which the global approach can provide opportunities for, but not limited to, student reading development.

1. Student involvement and participation in the internal and external governance of the institution. Not only student government activities, but involvement and membership on the governing board and major decision-making bodies of the institution. By accepting the responsibility, students find it necessary to prepare materials and employ language skills to keep abreast of current concerns.
2. In a true community situation students can become directly involved with the community and actively pursue and meet expressed needs. For example, a class or an individual can assist a family in the construction of a dwelling. Skills related to planning, architectual drawing, development of building specifications, and actual construction are required. Numerous other community-oriented projects can be implemented with meaningful student involvement.
3. Local political, economical, social, and ecological issues are areas in which student interest is high. Preparing testimonial materials and position papers, lobbying, and demonstrating require reading skills.

The list is incomplete and is limited only by imagination and student willingness to participate. Indian arts and crafts, student publications, curriculum development, and staff development are additional examples that can be employed.

In summation, the reading habits and behaviors of the American Indian college student are interwoven into the individual's way of life, language, values, and attitudes. Tribal and individual differences further complicate the matter. The total situation must be recognized and accepted before and educational program can effectively benefit the individual Indian student. The Indian community college, with its global concern, can provide a program that treats the Indian culture as equal to the non-Indian culture, resulting in self-awareness, which in turn influences student achievement.

References

Artichoker, J., Jr., & Palmer, N. M. *The Sioux Indian goes to college*. Pierre, S.D.: State Department of Public Instruction, 1959.

Bass, W. P. *An analysis of academic achievement of Indian high school students in federal and public schools.* Albuquerque, N. M.: Southwestern Cooperative Education Laboratory, 1971.

Jensen, K. D., & Jensen, S. Factors in college education for Indian students. *Improving College and University Teaching.* 1970.

McGrath, G. D. *Higher education of Southwestern Indians with reference to success an failure: Supplement*. Tempe: Arizona State University, 1962.

The Navajo Tribe, Division of Education. *Strengthening Navajo education.* Albuquerque, N. M.: Modern Printing Company, 1973.

Spencer, P. L. Tradition unmasked. In M. P. Douglass (Ed.), *Reading in education: a broader view.* Columbus, Ohio: Charles E. Merrill, 1973.

U. S., Commission on Civil Rights. Socio-economic profile of American Indians in Arizona and New Mexico. Unpublished paper, Washington, D.C., 1972.

U. S., Congress, Senate, Committee on Labor and Public Welfare. Special Subcommittee on Indian Education. *Indian education: A national tragedy – A national challenge.* 91st Cong., 1969, S. Rept., 1969.

INCREASING PHONICS PROFICIENCY OF PRE-SERVICE TEACHERS THROUGH MODULARIZED INSTRUCTION

HENRY D. OLSEN
Medgar Evers College

ABSTRACT

The purpose of this study was to identify the level of expertise in phonics analysis of prospective teachers, and to determine as to whether instructional modules could be developed, instructional materials accumulated and/or organized, to teach phonics to prospective elementary school teachers at an 80% minimum level of competency.

The *Phonics For Teachers Test* was administered to 52 prospective urban teachers to determine as to whether they had sufficient or insufficient knowledge of phonics, if the student had sufficient knowledge of phonics then the student could exit from the module entitled "Linguistic Proficiency." If the student had sufficient knowledge of phonics then the student was assigned the module. Upon completion of the module assignments the student was administered the Phonics For Teachers Test in a rearranged order as a post test.

Results of the pre-test identified that 51 of 52 students had to complete the instructional module entitled "Linguistic Proficiency." The scores on the eight subtests of the Phonics For Teachers Test ranged from 57% to 78%. As a result of completing the instructional module the eight subtest scores ranged from 67% to 89%. Applying the sign test to the total score change from pre-test to post test indicated that the changes in scores were significant at the .001 level.

The study did identify that modular instruction could increase a prospective teacher's knowledge of phonics and phonics principals. However, it was stressed that further studies must be udertaken to insure that competency in phonics by a teacher also produces better phonics skills in the children taught.

A CRITICAL EVALUATION OF SEVERAL METHODS FOR MEASURING GAIN IN READING

EARL F. RANKIN
University of Kentucky

LOWELL EBERWEIN
University of Kentucky

Upon learning of the "residual gain" technique developed by Manning and DuBois (1962), Rankin published two papers (1965, 1967) in which the superiority of this technique over simple measures of "crude gain" was demonstrated as applied to reading test scores. Davis (1961, 1972) published two important papers in National Reading Conference yearbooks in which he described several different methods of measuring changes in reading. These papers stimulated considerable interest in this topic. The measurement of reading improvement is understandably a matter of great concern to both teachers and researchers of reading. However, attempts to measure reading gains are so fraught with difficulties that there is a need to take another look at the assets and liabilities of various measurement procedures and the proper use of these procedures in the measurement of reading improvement.

General Issues in Gain Measurement

For sake of brevity, a number of issues related to all types of gain measurements will be listed. Although these issues have been discussed in several books and articles on measurement, knowledge of these matters is not widely disseminated.

1. The test used in selecting poor readers (or superior readers) for gain measurement should not serve as the pretest for use in measuring their improvement.
2. Standards of test reliability are higher for gain measurements than for individual measurements due to the presence of error in both measurements involved in estimating the true gain score. To attain this objective, the administration of two pretests and two posttests is desirable.
3. It is important that both pre and posttests be sufficiently difficult that no one receives a perfect score and sufficiently easy that no one misses all the items.
4. If the same test is not used before and after treatment, parallel forms should be used which have been rigorously equated.
5. Factorially pure tests (or subtests) best satisfy an important

assumption, in measuring gains, that two individuals with the same true score have the same amount of ability on a given trait. A test of a carefully defined specific reading skill meets this requirement better than a general reading comprehension test.

6. Simple differences between pre and posttest measurements (i.e., crude gain) require that measurements be expressed in interval scales with identical zero points. Other types of gains expressed in relation to regression lines require that results be expressed on a common reference scale such as a standard score scale. In comparing gains between individuals starting from widely separate scores on an initial test, it should be assumed that equal intervals throughout the scale represent equal increments of ability. Since few, if any, reading tests strictly satisfy this criterion, more confidence can be placed in the comparison of gains between individuals starting from the same, or similar, test scores than between individuals having different initial test scores.

7. The higher the correlation between pre and posttest scores, the greater the gain must be in order to exceed the error of measurement involved in both measurements. The lower the correlation between tests, the higher the reliability of both tests must be in order to achieve reliable gain measures. However increasing reliability of pre and posttests tends to increase the correlation between them.

8. It is best to have separate reliability estimates on both pre and posttests based on the data actually used in the gain measurement itself. In the absence of such data, publishers' reliability coefficients should only be used if modified appropriately to adjust for differences in the range of talent between the standardization group and the sample involved in the gain measurements.

9. If gain is used as a variable in correlational studies, the effects of any relationship between the pretest and the other variable, W, must be partialed out of the obtained correlation. Otherwise, the correlation may reflect the relatioship between the pretest and W rather than between the gain and W.

Several of the previously mentioned criteria for using gain measurements are individually difficult to meet in practice. The cumulative effect of failing to meet several criteria should raise very serious questions about the accuracy of gain measurements.

Types of Gain Measurement

Four types of "change estimates" were described by Davis (1972). The term "gain" is used in this paper because of widespread usage. Actually, a "gain" is a change in either a positive or negative direction. Although gains are actually "estimated," the term "measurement" will be used for convenience. These types of gains will now be considered in relation to measurements for individuals only. Davis' symbol system will be slightly modified using more conventional statistical symbols of X for pretest and Y for posttest. Definitions of terms are given at the end of this paper. The first type of gain measurement is the "crude gain" (i.e., "raw score point estimate" in Davis' terminology). This is expressed by the formula:

$$G = Y - X$$

Paradoxically, this is the most often used and, at the same time, the poorest way to measure gain for individuals. No doubt many teachers evaluate reading improvement using this specious formula. (A variation of crude gain is called "percent gain." The latter measurement has all the deficiencies of crude gain and will not be considered in this peper). Crude gain is easy to compute and misleadingly simple. Its greatest weakness is the overestimation of true gains for low scoring students on the pretest and the underestimation of true gains for high scoring students on the pre-test. It also produces biased correlations between pretest scores and gains because of the presence of measurement errors on the X variable which is found in both terms of the correlation: $r_{X}\,(Y-X)$. Correlations between crude gains and other variables are of doubtful value. The use of crude gains is not recommended. This raises serious questions regarding the proper interpretation of pre and post criterion referenced test differences in prescriptive reading instruction.

A second type of gain measurement is called the "regressed gain" (i.e., "regressed score point estimate," to use Davis' terminology). This gain is expressed by the formula:

$$G = r_{YY}Y - r_{XX}X + \overline{Y} - \overline{X} - r_{YY}\overline{Y} + r_{XX}\overline{X}$$

The term "regressed" refers to the regression of the true score on the observed score. That is, the true score is estimated from the observed score by use of a regression coefficient which, in this case, is the reliability coefficient of the test. Thus, the true score for both X and Y is estimated, and the difference is expressed as a regressed gain. The additional elements in the formula are provided to make the mean of the regressed gain distribution equal to the mean of the same distribution of crude gain. This is a relatively simple formula which, in correcting for error in both X and Y, reduces the error of measurement for the gain and gives a better estimate of gains for extreme X scores than does crude gain. It is not a "base free" measurement. That is, it does not correlate zero with the pretest score. If, by chance, the reliability estimates are exaggerated, the true score of both X and Y will be inflated. It is important to remember that reliability "measurements" are really only "estimates."

The third type of gain measure is the estimate of "true gain" (i.e., "point estimate of true change," as designated by Davis). This formula is:

$$G = \frac{S_Y(r_{YY} - r^2_{XY}) - r_{XY}S_X(1 - r_{XX})}{S_Y(1 - r^2_{XY})}Y + \frac{r_{XY}S_Y(1 - r_{YY}) - S_X(r_{XX} - r^2_{XY})}{S_X(1 - r^2_{XY})}X$$

$$+ (\overline{Y} - \overline{X}) - \frac{S_Y(r_{YY} - r^2_{XY}) - r_{XY}S_X(1 - r_{XX})}{S_Y(1 - r^2_{XY})}\overline{Y} - \frac{r_{XY}S_Y(1 - r_{YY}) - S_X(r_{XX} - r^2_{XY})}{S_X(1 - r^2_{XY})}\overline{X}$$

This formula represents a slight modification of Lord's (1966) formula for estimating true change. This formula makes use of r_{XY} in addition to the two reliability coefficients used in the regressed gain measurement. True gain is estimated from observed scores by means of a multiple regression technique. This is, in general, a superior measurement providing a smaller measurement error than any other technique. As in regressed gains, the gains of individuals with extreme X scores are not misleading if one uses this gain measurement. Once the rather complete set of data for the formula is compiled, the computation of gains for every individual in the group is routine. This gain measurement also does not produce a base free measurement which is uncorrelated with X scores. Like regressed gain results, true gains should not ordinarily be correlated with an outside variable without partialing out the effects of X on the correlation.

Several properties of true gain measurements should be taken into consideration in interpreting results. Marks and Martin (1973), using computer simulated data, found that the error of measurement of the Lord formula (which is almost identical with Davis' formula) was strongly influenced by the correlation between the true pretest score X_T and the true gain. In other words, gain measurements with reasonably small errors of measurement are only found for those distributions where individuals who performed the best on the initial test also gained the most between pre and post-tests. The most reliable gain measures were found where the correlation approached unity and reliability for both tests was high. As a general rule, the authors suggested that test forms should only be employed. . . "with reliabilities in excess of .85 and especially so if the true gain–initial true score correlation is expected or found to be .70 or less" (1973). Contrary to this advice, the present writers suggest that the measurement of true gains not be limited to these conditions. Rather, the standard error of measurement should be carefully considered in interpreting results. It should be noted that similar studies have apparently not been made on other gain formulas. It is not known whether the estimate of true gain is any more or less subject to the influence of these variables than are other types of measurements.

The fourth type of gain measurement is called "residual gain" (i.e., "point estimate of deviation from predicted change," in Davis' terminology). Residual gains are based upon all the variance in Y that is not correlated with the variance in X. They are thus "base free." The correlation between residual gain and X is zero. The formula is:

$$G = Y - \bar{Y} - \frac{r_{XY}S_Y}{S_X}(X - \bar{X})$$

This formula is algebraically equivalent to the raw formula presented by Tracy and Rankin (1967). Unlike the previously described gain estimates, this technique removes all regression effects from the gain measurement. As such, it is designed to enable gains to be correlated with another variable without allowing any relationship between X and the other variable to confound the interpretation of the correlation. These gains indicate the extent to which an individual gained more or less than expected. Since the sum of squared deviations below and above the regression lines of Y on X is zero, the mean of a given distribution of residual gains is always zero. This may create problems for those who wish to use these measurements for various research purposes.

Residual gain estimates entail several problems. In removing all regression effects from gain measurements, both regression due to measurement errors and to real movement toward the mean are removed. As such, residual gain scores are not ideally suited for measuring reading improvement for specific individuals so much as they may be for correlating reading gains with other variables. Unfortunately, it has been shown that in producing a correlation of zero between residual gain socres and observed X scores, there still remains a correlation between gain measurements and the true scores for X. Lord (1958) and Bereiter (1962) have shown that due to the failure to control the correlation between residual gain and true X scores and other characteristics of partial correlations, a positive correlation will usually exist between residual gain scores and any third variable that is correlated with observed X scores. Therefore, any correlations between residual gain scores in reading and other variables should be corrected for attenuation on the X variable. However, as indicated in the final section, other techniques are available to accomplish this objective.

A formula for residual gain has been developed by Tucker, Damarin, and Messick (1966) and modified by Davis, which attempts to produce gains which will correlate zero with the true score for X. This formula is as follows:

$$G = Y - \overline{Y} - \frac{r_{XY}S_Y}{r_{XX}S_X}(X - \overline{X})$$

This formula is devised to correct the previously described defect in the conventional residual gain score. It should be noted that Davis found that this measure does produce a higher standard error than the conventional formula.

Empirical Data

To study several characteristics of the previously described gain measurements, the writers analyzed reading test scale scores for 258 fourth-grade children. Tests were administered to the same subjects one year apart. The following types of gains were measured: crude gain, regressed gain, true gain, and both of the previously described residual gains. The following results are of interest:

1. As expected, the mean gains were identical for crude gain, regressed gain, and true gain. The mean gain for both residual gain scores was zero.
2. Correlations were .99 among all combinations of gain scores. Although high correlations are usually found among gains, these correlations were inflated by the use of a spuriously high publisher's reliability coefficient in calculating several gain formulas. In any event, high correlations between gains, together with the equal means, might lead some investigators to conclude that all formulas yield essentially the smae results. This conclusion would not be justified.
3. Correlations between pretest scores and gains were on the order of $-.15$ ($p < .05$) for crude gain, regressed gain, and true gain, and zero for the two residual gains. When the same correlations were computed separately for top, middle, and lower third students on the pretest, slight negative correlations of about $-.22$ ($p < .05$) were found for all gains but conventional residual gains for both top and lower third groups. These results for both the total group and extreme thirds indicate that some true regression is occurring after the effect of measurement error is reduced. Even true scores tend to regress toward the mean. This statement is not an explanation; only a description. Causative factors in the real world must account for such phenomena. It should be noted that this "true regression" is covered up by the conventional residual gain technique.
4. Errors of measurement were in accord with theoretical

expectations arranged in descending order from crude gain to regressed gain and true gain. Measurement errors for the two residual gain formulas were similar to the standard error of measurement for the regressed gain but considerably higher than measurement error for true gain. When a confidence interval for measurement errors was determined by multiplying 1.96 x the standard error of the measurement, the error for all gains, with the exception of true gain, was substantial. For example, when grades were assigned to gain scores on a normal curve, the gain was only accurate to within a range of plus or minus one letter grade for all grades but true gain. True gains, however, were accurate within a range of slightly over one-half of a letter grade. These errors were based on a simulated reliability of .90 and pre-posttest correlation of .79. When a simulated reliability of .96 was used, the error of all gains was sufficiently small to be accurate for a given letter grade. In as much as these differences reflected errors across broad grade categories the interval within which the true gain might be found for individuals was substantial, with the exception of true gain. Thus, it can be seen that considerable measurement error exists for even a large sample of data collected on a test with a substantially high reliability coefficient.

Uses of Gain Scores for Studying Reading Improvement

Several of the following suggestions are based with some modifications, on an excellent article by Cronbach and Furby (1970). There are three types of questions about reading improvement that one might like to answer with gain scores. There are also three reading research purposes for which reading improvement scores might be desired. Suggestions will now be made concerning the proper uses of gain measurements and the selection of gains appropriate for particular purposes.

Question 1: Has a student improved in reading? If a student demonstrates gains that cannot be accounted for in terms of measurement error (i.e., 1.96 x standard error of the gain), it can be concluded that he has improved in reading skill. The best gain technique for this purpose is true gain, although regress gain scores could be used if the necessary information is not available for estimating true gain.

Question 2: Has one student improved in reading more (or less) than another student? As pointed out previously, this question can best be answered for students having similar or identical scores on the initial test. For this purpose, either regressed gain or true gain should be used. For comparing students having widely different initial test scores, it is desirable to have two factorially pure tests yielding scores on an interval scale where there is a substantial basis for assuming a one-to-one correspondence between score increments and ability increments at all scale levels. If this can be assumed, the true gain technique is most appropriate for the latter purpose.

Question 3. Has a student (s) improved in reading more or less than expected? For this purpose, some type of residual gain is suggested. As Cronbach and Furby (1970) have indicated, a conventional residual gain score will select the same individuals for this purpose as in more sophisticated formulas, such as a residual true gain based on Lord's formula (see Cronbach & Furby Formula 23). However, if further analysis of the gain data is desired, the use of a residualized true gain formula, such as Davis' modified residual gain of Cronbach's Formula 23, is suggested.

Research Purpose 1: Should gain measurements be used as a dependent variable in a reading research design? There is no need to use gain for this purpose. As Cronbach and Furby (1970) pointed out, an analysis of covariance will serve the same purpose with much less risk.

Research Purpose 2: Should gain measurements be used as a criterion variable in a correlation study? Again, there is no need for gain measurements in such studies. The use of partial correlations, with proper corrections for attenuation of the zero order correlations entering into the partial correlation, will accomplish this objective without the attendant difficulties of gain measurement.

Research Purpose 3: Should gain measurements be used to measure some reading construct (e.g., "reading as a learning process")? If so, this might entail measuring gains between information possessed before and after reading. This is an intriguing prospect, but it may be easier to attain this objective through test construction rather than through gain measurements with all of their difficulties. The use of gain measurements for this purpose would entail the difficult task of comparing individuals with widely different initial reading skills and meeting the difficult assumptions involved in such an undertaking. If this were possible, Cronbach and Furby's (1970) modification of Lord's true gain formula to allow for the estimate of gain based upon more than one pretest variable and several posttest variables might be a fruitful strategy.

Contrary to recommendations in previous papers by Rankin (1965, 1967), it is suggested that conventional residual gains not be used for estimating reading improvement of individual readers unless data is not available for calculating several previously recommended alternate gain formulas. However, conventional residual gains are still preferable to the use of crude gains for this particular purpose. In addition, conventional residual gain is a valid measurement for use in selecting readers who have gained less (or more) than expected for treatment or research purposes.

Definitions of Symbols

X = individual score on pretest

Y = individual score on posttest

$\overline{X}$ = mean score on pretest

$\overline{Y}$ = mean score on posttest

S = standard deviation

r_{XX} = reliability coefficient on pretest

r_{YY} = reliability coefficient on posttest

r_{XY} = correlation between pretest and posttest

References

Bereiter, C. Some persisting dilemmas in the measurement of change. In Chester W. Harris (Ed.), *Problems in Measuring Change.* Madison: University of Wisconsin Press, 1963.

Cronbach, L.J. & Furby, L. How we should measure "change" – or should we? *Psychological Bulletin,* 1970, *74,* 68-80.

Davis, F.B. The assessment of change. In E.P. Bliesmer and A.J. Kingston, Jr. (Eds.) *Phases of College and Other Adult Reading Programs.* National Reading Conference 10th Yearbook, Milwaukee: The Conference, 1961.

Davis, F.B. A comparison of statistical models used in the measurement of change. In Frank P. Greene (Ed.) *Investigations Relating to Mature Reading.* National Reading Conference 21st Yearbook, Milwaukee: The Conference, 1972.

Lord, F.M. The measurement of growth, *Educational and Psychological Measurement,* 1956, *16,* 421-437.

Lord, F.M. Further problems in the measurement of growth. *Educational and Psychological Measurement,* 1958, *18,* 437-451.

Manning, W.H. & DuBois, P.H. Correlational methods in research on human learning. *Perceptual and Motor Skills,* 1962, *15,* 287-321.

Marks, E. & Martin, C.G. Further comments relating to the measurement of change. *American Educational Research Journal,* Summer, 1973, *10,* 179-191.

Rankin, E.F. Jr. & Tracy, R. J. Residual gain as a measure of individual differences in reading improvement. *Journal of Reading,* 1965, *8,* 224-233.

Tracy, R. J. & Rankin, E.F. Jr. Methods of computing and evaluating residual gain scores in the reading program. *Journal of Reading,* 1967, *10,* 363-371.

READING AND OTHER LANGUAGE SKILLS OF RESIDENTS IN A STATE SCHOOL AND HOSPITAL

EMERY P. BLIESMER
Pennsylvania State University

VINCENT S. ORLANDO
Pennsylvania State University

JOSEPH A. MUIA
Madison College

For approximately the past 10 years, the Special Education Department of the Pennsylvania State University has carried on a funded project at a state school and hospital for mentally retarded. This has mainly involved conducting a demonstration school for school age (up to21 years) educable and trainable residents. Penn State's direct role in the project was terminated last June, 1975; but it is presently being continued, in a modified way, by an area public school district with which the project was conducted cooperatively the last several years. University staff members involved in the project were mainly members of the division of Education of Exceptional Children of the Department of Special Education; but the role of reading and language development consultant was filled by the first presenter of this present paper, a member of the division of Developmental and Remedial Reading, and a succession of doctoral graduate assistants, who worked with teachers and students in the demonstration school, and some of the other residents also, an average of approximately one day a week. The last two assistants are co-presenters of this paper. The first assistant, Dr. Johanna Strollo, later served as principal of the demonstration school, presented a paper on her dissertation at this conference three years ago, and for the past two years has been assistant superintendent of the state school and hospital involved.

The several years experience with the demonstration school project resulted not only in benefits and advantages for residents of the state school and hospital; the work and observation of University staff members in the project also resulted in considerable learning on the part of these participants. Several of the major observations will be treated briefly in the following pages, under the headings indicated.

Measuring Achievement and Competency Levels

Working with the various types and degrees of retarded, handicapped individuals made it necessary to modify many typical testing and evaluation procedures, which yielded some surprising and helpful information. One woman who was in a relatively reclining position most of the time, and could use only one hand with great difficulty, was usually considered to be pointing wrongly because it was found she could not make her finger point to the place she intended. By the time she was able to move her finger to another position, the person working with her would already have moved on to another activity or question. Consequently, she was seldom able to indicate that she could respond correctly. Learning to wait for her responses and to check on whether these were at the spot she desired led to discovering much higher potential and competency levels than had been previously anticipated. On a typical reading readiness test, for example, she was found able to respond successfully to many of the items whereas previously she had been judged as performing successfully on relatively few. This woman could also move parts of her body with only great difficulty, except for relative skill developed with one foot. After a time she not only learned to weave with a large needle with this foot but also to pass a thread in and out of material on both sides and do designs with needle and thread with her foot. She also became quite adept at using her foot for "talking;" and those who worked with her also became quite adept at understanding her "foot talk."

A method of having a clinician point to possibilities for responses and having the examinee react with body indications, such as blinking, winking, shaking of head, and so forth, was also improvised on further. Use of ability to wiggle a big toe and a button-pusing procedure are alluded to in other sections of this paper.

Continued work with given groups of students and some older residents led to surprising findings with regard to levels of skill and competency possessed by some residents. A number of upper level students were found able to score at second through sixth or seventh grade levels on standardized reading and other tests and at fifth to seventh grade levels on spelling tests. A somewhat similar situation with regard to achievement or competency status was found with a number of residents who were not in an academic program but were in a work-study program for older residents (ages 20 to approximately 50).

A number of residents were found to have developed a surprisingly considerable degree of phonetic analysis skill. This is especially surprising when one considers that phonetic analysis is basically a matter of generalizing about sounds in words that are known and applying these generalizations to words not known or recognized immediately, and considers further that generalizing skill is supposedly one of the major weaknesses in the mentally slow or retarded. There is suggested a definite need to look further into just what enables mentally slow individuals to attain this degree of phonics skill and to reassess the role of rote learning, memorization, and other drill procedures which are often regarded negatively from a pedagogical viewpoint.

A desire to read for enjoyment or pleasure purposes was expressed by some older residents. One 37 year-old woman with approximately third grade level reading skill expressed a desire to read about "kitty cats, rabbits, and puppy dogs;" but she could not be provided with sufficient and appropriate supervision and guidance to enable her to follow through with this. A thirty-year-old male who scored at approximately a seventh grade level on a slight word test and had approximately a fourth grade reading comprehension level, indicated that he liked to read *Weekly Readers* and that he had a "whole pile stored in my locker." Exploratory work carried on by an assistant during a recent summer term indicated that regular and direct supervision and guidance in the area of reading for information and enjoyment could have some very definite results and benefits for some residents.

Much interest has been shown in recent years in "dyslexia," "learning disabilities," and similar areas with other labels. With this interest there has also often been accompanying doubt, confusion, and skepticism. In working with residents of types found in the project, one soon learns not to judge overall or general lack because of an apparent deficiency or lack in one or two areas. Several non-ambulatory residents were found to have remarkable vocabulary and speaking skill, talking somewhat like upper high school or lower college level students. However, evaluation and analyses of various types indicated a complete lack of reading skill, in fact, a complete lack of ability to identify name, or discriminate visually among printed letter forms. Another multiply handicapped individual, in his late teens, after a number of years of being regarded as merely existing, was discovered to have a considerable degree of understanding of conversation and other events about him and appeared actually to have developed some degree of reading skill. However, it was extremely difficult for him to indicate or show responses or understanding. (This was the individual with whom the wiggling of the big toe was begun to be used for indication of a positive response. Efforts to pursue this means of responding further and to use the toe wiggle to trigger other mechanical or electronic responses were, unfortunately, terminated by the young man's sudden unexpected departure from the institutional setting.)

Reference Points to go Back To

Basic to some pedagogical principles is the opportunity to go back to previous learning points if mastery at a certain level has evaded or escaped one, that is, having something to fall back on if one forgets. One class worked daily in small groups and made use of a language experience type approach in dictating news of the day for the teacher to record on the blackboard. Such common elements as the day of the week, the date, words describing the weather, and words for familiar items of clothing regularly appeared in the daily news. Listing of days of the week and of the months, charts with pictures and corresponding weather or clothing words, and other similar types of charts were posted near the groups in the classroom. Those individuals who forgot how to spell a word or could not remember what the particular symbol for a day or an item of clothing said could be referred to the appropriate chart and get the spelling or recognize the word by, for example, counting from the first word listed or reciting the days of the week in order (which they could usually do) until they came to the day in question and, thusly, get the day and/or the right spelling. Similar procedures were utilized with regard to the letters of the alphabet, which were arranged in order on wall cards where they could readily be seen. Number charts were also used in this way. While this idea was not a very original or new one, the need for giving multiply handicapped youngsters such aids or reference points, and for leading them back to such points again and again, has not always been recognized and considered.

Similar procedures were found to be helpful with older pre-vocational groups, especially with pictures and labels for tools, daily routine duties, types of weaving stitches, and the like. To avoid having lables of words be memorized as positions or specifics on a wall or chart, frequent shifts or changes in position of labels and illustrations needed to be made, with students frequently needing to be led back to earlier reference points.

Use of Language Boards

Observation of the use of "language boards" by speech therapists with some of the multiply handicapped residents suggested more direct and intensive use of such boards for developing reading and other communication skills with subjects who had extreme difficulty in indicating needs and other responses. Basically, the language board is a device set up or arranged so that a subject can indicate a need or response by saying, pointing to, or otherwise designating a printed word or words, and object (s), a picture (s), or a category or categories listed or arranged on a board, with the extent of the listing of words or categories depending upon the level of mental and/or other development of the individual. As a greater number of ways of indicating responses was discovered or developed a number of subjects were noted to have greater potential than had been estimated previously, Among points observed and suggestions made which led to some changes in the practices of speech therapists and reading and other teachers were the use of lower case, manuscript letters instead of capital and cursive letters; labeling items in various lists by letters or numbers and then having individuals indicate a letter or number of response rather than try to point or indicate the response on a particular place on the board; pointing to a board item and having a subject indicate a negative or positive response with a wink, a blink, a slight head shaking, a grimace or other motor indicatsion. With one young man, for whom hand and other voluntary body movements were extremely difficult, the ability to wiggle the big toe of one foot fairly readily was begun to be used as his indication of a positive response. Another resident, operated a device which enabled her to push a button and rotate an arrow until it reached a numbered or lettered area which indicated the desired response category. Some individuals could, after a time, build sentences or partial sentences by indicating responses as described, with improvement being noted in such as time went on. Another severely handicapped woman was reported to be acquiring a slight vocabulary of over 100 words after several months of suggested work with a language board. The language board still offers many further, fruitful, exploration possibilities.

Influence of Foster Grandparents

One of the especially noted aspects relative to the project was the use of retired or senior citizens in a funded "foster grandparent" program. Coming from the surrounding area, each grandparent was assigned one (usually) resident to attend to and to work with under supervision and guidance, for four to five hours a day, for four to five days each week, working on various language and/or motor skills.

The observed effect of this sustained, patient, and apparently loving attention and work on a number of profoundly retarded subjects over a period of time was especially remarkable; but the influence on a number of educable level residents with various learning disabilities and multiple handicaps was particularly noticeable also. One demonstration school girl with a number of physical and mental handicaps had developed a noticeable sight vocabulary and oral spelling skill after having worked over a year with an eighty-one-year-old grandmother and was also exhibiting some improvement handwriting skill. A retired college professor had helped a boy with muitiple physical and mental handicaps develop a noticeable skill in mkaing changes with actual coins and recognizing some words. This grandfather had also succeeded in getting another severely handicapped boy to succeed in pushing his wheelchair up and down a hallway and into classroom with the aid of his one usable foot, as well as getting the boy to make more definite and recognized responses to materials written on a blackboard. Improved use of language boards by several residents was also noted after guided and persistent attention by foster grandparents.

A variety of procedures was found to be used by grandparents, including the spelling out of words to teach sight vocabulary, memorizing, imitating, repetition without changing context or setting, and so forth. While trained educationists might regard some of the procedures used somewhat negatively and in violation of some basic learning principles, mastery of skills achieved by residents under grandparents' care suggests a need for reanalysis and further look at some of the procedures, especially with respect to particular effects of long-term attention with regular and consistent patience, oral training methods, use of extrinsic rewards, and the like (particularly with regard to the development of basic sight word vocabularly, lower level word attack skills, and basic arithmetic skills).

Influence of the program on the grandparents themselves also was quite noticeable. This suggests possibilities for making retirement less frustrating and fearful for senior citizens in our society by utilizing our older citizens in comparable roles in other than federal or special funded programs as well.

Summary

The foregoing pages represent a brief, selected summary of some of the major observations made and impressions obtained, after several years of work with a university demonstration school project in a state school and hospital relative to the status and development of reading and other language skills of state institution residents. These were briefly discussed under the following five general headings:

1. Measuring achievement and competency levels of residents.
2. Status of achievement and competency levels of residents.
3. Reference points to go back to.
4. Use of language boards.
5. Influence of "foster grandparents" program.

As was anticipated, considerable benefits, advantages, and learning resulted for institution residents as well as for University staff project participants; and considerably further learning and other benefits seem indicated for the future.

AN EXAMINATION OF STUDENT COPING STYLE, TEACHER CONTROL AND STUDENT ACHIEVEMENT IN READING

HOMER COKER
West Georgia College

JEFFREY LORENTZ
West Georgia College

ABSTRACT

Systematic classroom observation instruments have, in recent years, been refined to the point where serious investigators can use them in studies of teaching effectiveness. Two such instruments, Spaulding's CASES and Soar's FLACCS, are among the observation instruments used in the Carroll County CBTC Project. In the project, data have been collected on more than 1800 students in 60 on-going classrooms, grades 1 through 12 during the 1974–75 school year. Traditional student measures included standardized reading tests given to each student in the fall and again in the spring.

A number of recent studies have examined the relationship between observed classroom behaviors and student growth including those by Cobb (1972), Lahaderne (1968), and Samuels and Turnure (1974). These all found significant relationships between attentiveness and achievement. A modified version of CASES was used by Mckinney, Mason, Perkerson, and Clifford (1975) in a study which found that combinations of discrete classroom behaviors predicted achievement as well as IQ tests and that combinations of test data and behavioral information predicted school year achievement better than predicted by either alone. The Soars (1970–71, 1971, 1973) have found a number of significant relationships between pupil achievement and classroom behavior as well as between student achievement and influences outside the classroom such as socioeconomic status. They also found a number of non-linear relationships and interactions among the variables.

The instruments used in the present study included student reading achievement tests: The Scott Foresman Initial Survey Test and Comprehensive Test of Basic Skills (1st Grade) and the Iowa Tests of Basic Skills (Grades 2–7); observation instruments: CASES (administered pre and post) and FLACCS (year-long measure); and an estimate of each student's socioeconomic status.

The analytic procedure first involved the reduction of the data to classroom means, including pretest and posttest, SES, pre- and post-CASES (for teacher-directed and program-directed settings) and

FLACCS competency scores. These 49 variables for the 34 teachers were used in a series of stepwise multiple regression analyses to identify the best set of predictors of end-of-year reading achievement.

The results revealed that pretest and SES were major correlates of student achievement (accounting for a major portion of the variance) and that teacher control and student coping style were related to student achievement. The notion that the two observational instruments were interrelated was also supported.

SELF CONCEPT IN READING INSTRUCTION PROGRAMS

MERVIN D. LYNCH
Northeastern University

ANN MARIE HAASE
University of Massachusetts

Reading researchers have focused mainly on cognitive effects of reading instruction programs. However, reading instruction does not take place in a vacuum. Rather, an affective variable such as self concept may play an important role both as an input and output of reading instructional programs. Some studies such as that by Roth (1959) have shown that children with reading disabilities tend to have more defensive self concepts and this suggests that negative self concept changes will result when children show deficits in reading performance. Other studies such as that by Wattenberg and Clifford (1964) found that measures of self concept of kindergarten children were predictive of reading achievement at the end of second grade.

These and other studies have shown that self concept and reading achievement changes will affect each other interactively, but there is little evidence on what such an interactive pattern will be. The purpose in the present paper was to present the results of a series of evaluations of tutorial reading programs where both self concept and reading achievement measures were obtained. Three seperate studies are reported.

Method and Procedures

Study 1 was an evaluation of a summer tutorial reading program for 65 kindergarten children. Study 2 was a one year study of 46 children in a reading tutorial program. Study 3 was an evaluation of children during each of 3 years of a reading tutorial program. In each study pre- and posttest measures were obtained on indices of reading achievement and self concept for each of three years.

Measures of self concept were obtained with the children's semantic differential administered individually on a verbal basis. Ratings were obtained on 6 concepts. These were, (a) "How I feel about myself; (b) How I would like to be; (c) How I feel when reading by myself; (d) How my teacher feels about me; (e) How I feel about my homeroom teacher; and (f) How I feel about my reading teacher." Concept e was dropped from the kindergarten children and the concept "How I feel about my being ready to read" was substituted for concept c.

The D statistic between "How I feel about myself, and each of the 5 other concepts provided separate measures of 5 self concept variables. Global self concept measures of the idealized self concept (a vs. b) and teacher empathy (a vs. d) and situationally specific self concept measures of "personal security with reading alone (a vs. c), teacher identification (a vs. d) and reading teacher identification (a vs. e). A measure of self enhancement was obtained by comparing pre- and post-D statistic measures in each study. Finally, a measure of self worth was obtained using the sum of evaluative item scores on the concept "How I feel about myself."

Reading achievement scores were obtained on measures of word meaning and paragraph meaning on the Stanford Achievement Test in study 1 and with the American Achievement Test during studies 2 and 3. The kindergarten children in study 1 were tested with the Metropolitan Reading Test.

Results Study 1

Analyses of both reading achievement gains and changes in self concept were made with the correlated **t** test. Tests on self enhancement were made with the Wilcoxan Signed Ranks Test on the shifts in meaning shown on five D statistic scores. Sixty-five kindergarten children were given teaching in basic auditory and visual reading skills in the summer reading program. Results of the correlated **t** test showed that there were sizeable increases in teacher identification ("How I feel about myself" and "How I feel about my reading teacher), $t = 2.21$, $p < .05$, personal security with learning to read ("How I feel about myself" and "How I feel about being ready to learn to read"), $t = 2.21$, $p < .05$, and a tendency toward greater self idealization ("How I feel about myself" and "How I would like to be"), $t = 1.89$, $p < .08$.

These changes were accompanied by sizeable changes in reading readiness on World Meaning, $t = 2.06$, $p < .05$, Listening, $t = 2.36$, $p < .05$, Matching, $t = 8.90$, $p < .01$, and Knowledge of the Alphabet, $t = 7.25$, $p < .01$, sub-tests of the Methopolitan Reading Readiness Test.

Results Study 2

Forty-Six children in grades 2–6 were given reading tutorial work in small groups of 4 and 5, five days a week from January to May. There were no significant shifts in mean D values between the concept "How I feel about myself" and other concepts in the pre-post comparisons. When the mean Ds for all five concept pairs were compared together with a Wilcoxan Signed Ranks Test, there was a shift in meaning such that the concepts became more congruent with the judgment of the self concept ($p < .05$). These results indicated an increase in self-enhancement and this apparently was due to the tutorial experience. This increase in self-enhancement was accompanied by gains in reading achievement on means for both sub-tests of word meaning, $t = 6.19$, $p < .01$, and of paragraph meaning, $t = 7.09$, $p < .01$.

Results Study 3

In study 3 children in grades 4–6 were given reading tutorial work. Data was gathered for each of three years for 51 children in 1972–73, 48 children in 1973–74, and 49 children in 1974–75 school years.

There were no significant shifts during any of the three years for the mean D statistics nor on the mean self worth measure. However, when these compared overall, with the Wilcoxan Signed Ranks Test, there was a significant shift ($p < .05$) in self enhancement during each of the three years of the testing period.

This shift in self-enhancement was accompanied by significant gains in mean word meaning for 1972–73, $t = 5.98$ $p < .01$; 1973–74, $t = 6.45$, $p < .01$; and 1974–75, $t = 8.43$, $p < .01$. There were also significant shifts in mean paragraph meaning for 1972–73, $t = 7.12$, $p < .01$; 1973–74, $t = 2.58$, $p < .01$; and 1974–75, $t = 5.90$ $p < .01$.

While there were no significant shifts on mean D values for any of the 3 years, the end result of the three years of reading tutorial work showed a significant increase in self concept on all but one of the measures as indicated by the mean post test D values and mean post test self worth scores.

Conclusions and Implications

The results of these three studies indicate that changes in self concept do accompany changes in reading achievment. The effects of such programs on self concept measures appear to be small but accumulative. In this regard the only significant changes shown on global measures of self concept were for the kindergarten children in the summer remedial reading program. For the remainder of the studies significant mean changes were shown only when overall comparisons were made either across indices such as the use of the Wilcoxan test to obtain an indication of increased self-enhancement or when comparisons were made on mean post test scores taken over the three year period of time. These findings suggest that researchers who wish to investigate effects on global measures such as those for self concept should not plan to obtain large shifts in self concept for a single reading tutorial program. Rather the researcher should plan to use a variety of measures of self concept to obtain an overall index of self-enhancement. Or, possibly the researcher should plan to index global self concept effects of reading programs on a long range basis.

References

Lynch, M. D. *The D statistics as a multi-dimensional measure of meaning.* Paper presented at the Convention of the American Association for Research in Education, Chicago, April, 1972.

Lynch, M. D., Schacter, J. & Cochrane, T. *Development and validation of a set of semantic differential scales for measuring affective reactions of elementary grade level children.* Paper presented at the Convention of the American Association for Research in Education, Chicago, April, 1972.

Roth, R. M., The role of self concept in achievement. *Journal of Experimental Education,* 1959, 265–271.

Wattenberg, W. W., & Clifford, C. Relation of self concepts to beginning achievement in reading. *Child Development,* 1964, *35*, 461–467.

A LONGITUDINAL STUDY OF FACTORS WHICH ARE PREDICTIVE OF READING SUCCESS OF LOW SOCIO-ECONOMIC CHILDREN IN THE SEVENTH YEAR OF SCHOOL

LESLIE MCCLELLAN
University of South Florida

ABSTRACT

The purpose of the study was to identify those factors which are predictive of success in reading in the seventh year of school for children from a low socio-economic level. Included in the study is the determination of the relationship between certain readiness factors measured at the kindergarten and early first grade levels and achievement in reading from low socio-economic level children at the beginning of the seventh year of school.

This report is part of a longitudinal study of 422 children who were administered the Metropolitan Achievement Test (MAT) in the first grade and who had been administered the Metropolitan Readiness Test (MRT) Form A, the Peabody Picture Vocabulary Test (PPVT) Form A, the Goodenough-Harris Drawing Test, the MRT Form R, and the MAT during their kindergarten year. The kindergarten experience was a federally funded compensatory pre-primary education program specifically designed to bring about increased school readiness by helping each child develop increased self-esteem, language skills, and motivation for school success.

The subjects were administered the Metropolitan Achievement Test, Form G, at the beginning of the seventh year of school. Fifty-six of the children had been retained at some point in their school experience. Scores for the children were interpolated to seventh grade scores as indicated in the Metropolitan Test Manual. The dependent variables considered during the seventh year of school were scores on the MAT in word knowledge, reading comprehension, and total reading. The simple correlation .61 between the score in word discrimination on the MAT administered in May of the first grade was the best single predictor of success in word knowledge. This value was increased to .70 with the addition of other identified predictor variables.

The single best predictor of success in reading comprehension in the seventh year was scores obtained in word knowledge on the MAT in the first grade with a simple correlation of .42. The correlation was increased to .49 with the addition of the listening score on the February administration of the MRT, the total score on the July administration of the MRT, the alphabet score, the total score, and the copying score on the February administration of the MRT.

The best single predictor of total reading was the score in word discrimination obtained on the MAT in first grade with a simple correlation of .48. This correlation was further increase; to .59 by addition in the step-wise regression analysis of the following independent variables in the order given: the number score on the February MRT, the reading score on the first grade MAT, the PPVT administered in early kindergarten, the Goodenough-Harris Draw-A-Woman Test given in kindergarten, the matching score on the MRT given in February of the kindergarten year, the reading test score from the MRT given in September of the first grade, and the total score from the September MRT.

CONCEPTUAL DIFFICULTY OF READING MATERIAL: IDENTIFYING CONCEPTUAL CONSTRUCTS FOR PREDICTION

ROBERT J. TIERNEY
University of Arizona

JOSEPH L. VAUGHAN, JR.
University of Arizona

In an article entitled Vocabulary and Concept development, Langer (1967) suggested that vocabulary has a "direct" and "essential" relationship to concepts and that difficulty in reading comprehension stems from difficulty associating written symbols with the concepts they represent. If the validity of these notions can be assumed, then logically the concepts represented by words within a selection provide some measure of the readability of that selection. The problem arises: how does one measure these concepts? Specifically, how does one measure the conceptual difficulty of a reading passage?

Measuring the Conceptual Complexity of A Reading Passage

Providing somewhat of a theoretical basis for the development of a possible construct, Langer (1967) and Johnson (1975) suggested that the quality and quantity of associations provide a measure of one's knowledge of a concept. As Johnson stated, meaningfulness is dependent upon the quality and quantity of the reader's associations. Similarly, Langer suggested that knowledge of a concept is dependent upon the richness of associations which can be defined in terms of qualitative and quantitative considerations.

Prior to measuring knowledge of a concept, another problem arises: how does one identify concepts? While the existence of a relationship between vocabulary and concepts has been argued, studies by Thorndike (1917) and Nacke (1970) suggest that words within a single passage vary in their "potency." It would seem that individual words within a passage vary in their contribution to the conceptual complexity of a passage.

Given these guidelines, the authors attempted to identify those words which contributed to the conceptual complexity of a reading passage and thereupon, devise procedures to measure a student's knowledge of the concepts represented by these words. The following describes the procedures which were devised and implemented with a single reading passage, – an excerpt from "Born Free" by Joy Adamson.

1. *Identification of words contributing to the conceptual complexity* of a reading passage. Given the notion that words within a selection vary in their contribution to the conceptual complexity of a passage, only those words making a definite contribution to the conceptual complexity of the reading passage were selected. To this end, 12 teachers who were enrolled in a graduate reading course were given the task of identifying those words essential to an understanding of the selected reading passage. Those selected by the majority of teachers were considered as contributors to the conceptual complexity of the reading passage. For the purposes of the present study they were labelled conceptual identifiers.

2. Measurement of students' knowledge of concepts. Within the framework suggested by Langer (1967) and Johnson (1975), the quality and quantity of free associations for each conceptual identifier were used to measure the students' knowledge of the concepts required for the reading selection. A fluency score measured the quantity of free associations; specifically, fluency was the total number of responses to all stimulus words. To measure qualitative aspects, namely semantic and syntactic characteristics of the free associations, three measures were used. A flexibility score and commonality score assessed semantic aspects of the free associations, whereas a syntactic equivalency score assessed syntactic aspects. Specifically, flexibility was the number of definite categories of responses represented by the associations to the stimulus words. Commonality was measured by tallying for each student the number of times primary responses were elicited (primary responses being the three most frequent free associations elicited by the groups for each word.) Syntactic equivilency was the frequency with which the part of speech represented by the stimulus word corresponded to the parts of speech represented by the different free associations. Thus, considered simultaneously the four measures were assumed to measure a student's knowledge of the concepts required for an understanding of a reading selection.

Assessing the Validity of the Construct As a Measure of Conceptual Complexity

In order to examine the validity of the construct as a measure of the knowledge of the concepts of the reader, two groups that were expected to differ on the construct were compared.

The two groups were drawn from 247 eighth graders from a junior high school in Tucson, Arizona. The subjects were predominantly from white middle-class homes in suburban and semi-rural communities. Their performance on the S.R.A. Achievement Test, Reading Subtest, administered the previous year, indicated they were equivalent to the national sample used in the development of norms for that test.

During two class periods of approximately 20 and 60 minutes, two tasks were presented to these students. During the 20 minute period, students were presented with the free association task. The free association task required the students to respond to the five stimulus words isolated as conceptual identifiers for the selected reading passage. A time limit of one minute was applied to each stimulus word. During each one minute time period, the students wrote as many free associations that came to mind for each word. During the second period, a cloze passage was completed by each student. This passage was developed from the reading selection from which the stimulus words were selected. The passage was over 300 words in length and assessed as being at the eighth grade readability level by the Fry readability formula (Fry, 1968). With the first and last sentence intact, an every fifth word deletion schedule was used. Answers were recorded directly in the passage and exact replacements only were scored as correct.

To assess the validity of the construct as a measure of the conceptual complexity of reading material, the performance of subjects whose cloze scores fell among the upper and lower 27% of scores was compared on the four measures obtained for the free associations. Multivariate analysis of variance was used for testing the statistical significance of differences between the two groups. The computed multivariate **F** statistic for the combined vectors of students' associational scores produced a statistically significant coefficient, $F(4,125) = 9.25$, $p < .01$. Since the multivariate **F** was statistically significant, evidence for the convergent validity of the construct was provided.

In order to examine the contribution of the dependent variables to the difference between factor level means, both univariate F tests and the correlations between the discriminant functions and the original variables were consulted. The correlations and univariate **F**s are reported in the following table. While flexibility and syntactic equivalency contributed strongly, it is obvious by the size of the **F** values and relative value of the structured correlation coefficients that the variables taken together, and not separately, contributed to the differentiation between the two groups. The negative direction of flexibility's contribution does, however, raise an important related issue. It would seem logical that flexibility might contribute positively to comprehension, but not as measured by exact replacement on cloze. Perhpas the nature of these contributions is dependent upon the method of scoring cloze used in the present study.

Univariate F-tests and Structured Correlation Coefficiency

Variable	Univariate F (d.f. 1, 128)	Structured Correlation Coefficients
Fluency	1.318	.1865
Flexibility	3.847	-.3186
Commonality	2.303	.2465
Syntact. - Equiv.	5.313*	.3744

* Significant at .05 level

In order to assess the extent to which the present construct accounted for reading comprehension performance, a multiple regression analysis based upon the performance of all 247 subjects was undertaken. A multiple R of .33 ($p < .01$) was derived from this analysis. Thus, while the present construct accounts for a significant proportion of the variance associated with reading comprehension, much of the variance must be assumed to be derived from either other factors not associated with the present construct or factors related to conceptual difficulty not measured by the present construct.

An examination of the means and standard deviations indicated that the score distributions for the present population were all positively skewed. Apart from possibly understanding the extent of differentiation provided by the four measures, the effect of skewedness upon the multivariate analysis would be minimal (Myers, 1972). Ideally, further analysis of the present construct based upon a population exhibiting more normal score distributions would seem available.

Conclusions

The study was based on three assumptions; namely (a) that cloze was an appropriate measure of comprehension; (b) that the tasks used to approximate a student's competence (as opposed to performance) were appropriate, and (c) that the present procedures for the assessment of convergent validity were adequate. Relative to these assumptions, the present study has identified a method of differentiating comprehensions based upon a student's conceptual awareness.

References

Fry, E.B. A readability formula that saves time. *Journal of Reading*, April 1968, 11, 513-516, 575-578.

Johnson, R.E. Meaning in complex learning. *Review of Educational Research*, Summer 1975, 10, 3, 425-460.

Langer, J.H. Vocabulary and concept development *Journal of Reading*, 1967, 10, 443, 453, 455-56.

Myers, J.L. *Fundamentals of Experimental Design. Second Edition.* Allyn and Bacon, 1972.

Nacke, P.L. Skimming strategy in reading as a function of familiarity with content and redundancy reduction in printed discourse. Unpublished doctoral dissertation, University of British Columbia, 1970.

Thorndike, E.L. Reading as reasoning: A study of mistakes in paragraph reading. *Journal of Educational Psychology*, 1917.

SOME CONCURRENT VALIDITY STUDIES FOR GROUP TESTS OF WORD IDENTIFICATION SKILLS

P. DAVID PEARSON
University of Minnesota

BRIAN WEEKS
University of Minnesota

BEULAH NUVAL
University of Minnesota

In the press for efficiency of time and effort, the reading profession has recently developed a host of group tests of word identification skills. The constraints of the group testing situation require that these skills be tested as recognition rather than production tasks. Since the ability to apply word identification skills is the terminal objective of word identification instruction, recognition tests are valid only to the degree that they accurately predict what a student's performance would be on a production task.

The purpose of the set of studies reported in this paper is to assess the degree to which scores on group (recognition) tests approximate scores on individual (production) tests.

For example, does a consonant blend test which utilized the format in (1) provide information similar to that in production tasks like those represented in (2), (3), and (4)?

(1) *picture of a blimp* *(a) bl* *directions: circle the blend that starts the word represented in the picture.*
(b) pl
(c) cl
(d) br

(2) *bl* *[directions: can you say the sound of these letters?]*

(3) *blimp* *[directions: This word rhymes with imp; can you read it?]*

(4) *The black horse was blind.* *[directions: read this sentence]*
(Tester records whether or not bl blends are pronounced correctly.)

If the issues under investigation were simply measurement problems, then this line of research would probably not merit our immediate concern. Unfortunately, the more serious problems relate to day-to-day decisions that schools make concerning how students spend their time and energy during reading classes. If a student scores high on a group test of consonant blends, he may receive no more instruction on that skill. If he scores low, lots of practice may be prescribed. Hence the validity of our measurements is a practical and ethical as well as a technical problem.

The opening sentence of the paper asserted that group recognition tests of word identification skills are commonly used. Just to eliminate the possibility that we had set up "a straw man", we examined several basal reader testing programs, skills management systems, and group diagnostic tests. We have no straw man. In our survey, we found 15 distinct group formats, each with 2 to 5 minor variations. We found no production tasks. In addition, we found that many informal inventories (utilizing both group and individual subtests) use recognition tasks to assess phonics skills (e.g., Botel, 1961; Potter & Rae, 1973).

Background

The authors independently exhausted the previous research attempting to locate similar concurrent validity data and found one reference on a technical manual. The Stanford Diagnostic Reading Test (Karlsen, Madden & Gardner, 1966) contains a syllable recognition test (select the first syllable of the word you see) which correlated highly (r = .85) with students' ability to divide words into syllables. Unfortunately, they went only halfway since it would have been more informative to know the correlation between that subtest and the ability to pronounce multisyllabic words.

We found two descriptive studies which touched on the issue. Ramsey and Harrod (1969) administered one group test (The McKee Inventory of Phonics Skills) and two individual informal tests (a nonsense word pronunciation test and a substitution test – given **day** as a stimulus, respond with "say, may, lay, etc."). Teacher judgment were used to evaluate the efficacy of the measures. The teachers concluded that (a) the group tests lacked the accuracy to pinpoint weaknesses, (b) the nonsense test was useless because of real word interference, and (c) the substitution test caught more of the actual problems than did the other tests. Ramsey and Harrod did not examine statistical relations among the tests. Furthermore, it is likely that the tests did not measure identical skills.

Stevenson and Johnson (1974), in comparing children's ability to spell and pronounce nonsense words, found that their subjects' spellings differed from the patterns they saw when they pronounced the word 50% of the time. For example, a student might pronounce **bame**/bām/ but spell /bām/ **baym** or **baim.** Furthermore they found that children were more accurate in pronouncing than spelling nonsense words (accepting several alternative spellings as correct). The Johnson and Stevenson study is interesting in its own right but sheds little light on the present problem. People rarely use spelling measures as a substitute for pronunciation measures.

Experiment 1 – Vocabulary Recognition Tests

Purpose

The most common form of vocabulary assessment in basal reader and standardized tests is represented in (5); the second most common in (6)

(5) *teacher preacher teaching lasting*
[directions: circle the word I say] or
[circle the word for the picture at the left]

(6) *We have a new ______________ in school.*
teacher preacher teaching lasting
[directions: circle the word that best fits the sentence.]

Presumably, these group test formats are capable of assessing a student's ability to pronounce critical vocabulary items in isolation or context, respectively. Experiment 1 evaluated the degree of congruence between selected group and individual testing formats.

Method

Subjects

The sample for the study was drawn from among 4 second-grade classrooms in an average socio-economic level school in a Twin Cities suburb. All students who had completed Ginn 360 level 7 and whose parents permitted participation were used in the experiment (N = 40).

Materials

In order to obtain test items with appropriate difficulty levels, a sample of new vocabulary was drawn from levels 8 and 9 of Ginn 360. Half the words were judged to be **basic** (high utility but not decodable given previously taught skills) and half, **decodable** (given previously taught skills). From the population of available words, 24 were selected for inclusion in test format 1, another 24, for test format 2.

Test format 1 was the multiple choice format represented in (5). Distractors were generated by selecting 3 new words from the same levels which exhibited maximal available graphic similarity to the target word. For example, when **breath** was the target word, **break, drank** and **breeze** were the distractors. The students' task was to select the word the teacher said.

Test format 2 was the typical fill-in-the-blank format represented in (6). The words were different from those selected for format 1. Distractors were generated as for format 1.

Test formats 3 and 4 were individual formats. In format 3 the 24 target words used in format 1 were randomly arranged in a list of two columns of 12 words each. In format 4 the 24 target words used in format 2 were embedded into a passage of 110 words especially written at a third grade readability level in such a way as to accommodate the 24 target words.

Procedures

Each of the 40 students were randomly assigned to one of 4 test orders: 1234, 1243, 2134, and 2143. The group tests were always administered first. Appropriate instructions and practice items were given to all groups or individuals prior to test administration.

Scoring

A correct answer on test 1 or 2 was the selection of the target word. A correct response on test 3 or 4 required the correct pronunciation of the word.

Statistical Analyses

First, the overall repeated measurements ANOVA on test form was eliminated in order to allow for simple t test contrasts – test 1 versus test 3 and test 2 versus test 4. Number of errors per test was used as the dependent variable.

Second, correlation coefficients were computed to measure the relationship of test 1 to test 3 and test 2 to test 4.

Finally, a mastery level of 75% (6 errors) was established. Appropriate lines were drawn on a scatterplot of the bivariate distribution of sample scores, so that the degree of congruence of group versus individual "mastery" decisions was assessed by comparing the frequency of matches and mismatches on the two tests. Quadrants B and C represent matches; that is, cases in which the mastery decision for an individual are the same on either test. Quadrant A represents mismatches in which the test 1 decision is mastery and the test 2 decision is no mastery; Quadrant D, test 1 = mastery and test 2 = no mastery. This analysis was conducted in order to assess the practical decision making implications of obtained correlation and coefficients.

Results and Discussion

The mean number of errors, standard deviation, t values, p values, and correlation coefficients are reported in Table 1. Students made significantly more errors on individual test 3($\bar{X} = 5.35$) than on group test 1 ($\bar{X} = 1.35$), $t_{39} = -7.40$, $p < .001$. Recognizing, from among a set of distractors, a word said by the teacher was easier than pronouncing those same words on an isolated list. However students did not make all that many more errors on individual test 4 ($\bar{X} = 4.3$) than they did on group test 2 ($\bar{X} = 3.775$), $t_{39} = 1.70$, $p < .096$. Apparently selecting a word to fit a sentence context is not all that much easier than pronouncing those words correctly in context.

Table 1
Mean Errors and Standard Deviations
for Test Formats 1-4 (N = 40)

Format	1	3	2	4
$\bar{X}$	1.35	5.350	3.775	4.300
sd	2.237	4.475	4.135	4.536
	$t_{39} = -7.40, p < .001$		$t_{39} = 1.70, p < .096$	
	$r = .666$		$r = .903$	

The correlation coefficients reflect the relative similarity of scores on tests 2 and 4 (r = .903) compared to 1 and 3 (r = .666), recognizing that to some extent the low correlation of 1 with 3 may be due to the relatively smaller amount of variability on test 1.

The results of the scatter plot analysis of matches and mismatches are summarized in Table 2. In comparing group test 1 to individual test 3 using a 75% mastery level, there are 31 matches and 9 mismatches. However, all the mismatches are ones for which the group decision was mastery while the individual test decision was no mastery. Twenty-two percent of the time the group test decision is in error if one accepts that individual test as a legitimate criterion measure of vocabulary knowledge.

Table 2
Summary of Matches and Mismatches for Mastery Decisions

Test Formats 1 versus 3

Mastery Level		75% (6 or fewer errors)			83% (4 or fewer errors)	
Test 1 (group) Mastery	Yes	9	28	Yes	14	22
	No	3	0	No	4	0
		No	Yes		No	Yes
		Test 3 (individual) Mastery			(3)	

Test Formats 2 versus 4

Mastery Level		75% (6 or fewer errors)		83% (4 or fewer errors)	
Test 2 (group) Mastery	Yes	1	29	3	25
	No	10	0	11	1
		No	Yes	No	Yes
		Test 4 (individual) Mastery		(4)	

The data for tests 2 versus 4 are quite different. At the 75% mastery level there is only one mismatch (2.5% error), a finding easily predictable by chance. Even at a stricter 83% mastery level there are only 4 mismatches: 3 in which the group test decision is mastery but the individual test decision is no mastery and 1 in the opposite direction.

In comparing the data from the test 1 versus 3 comparisons with those of 2 versus 4, we are tempted to conclude that a group fill-in-the-bland test is a better substitute for its production test counterpart.

Experiment 2 – Consonant Blends and Digraphs

Method

Subjects

First-grade students (N=31) were randomly selected from among the 4 "average" reading groups (reading in Ginn 360 level 5) in a suburban elementary school located in a district whose population is predominately white middle class.

Materials

In all, 13 consonant blends (bl, tr, dr, spr, scr, pl, gr, sk, st, pr, cl, br, and gl) and 3 digraphs (wh, th, and ch) were tested in 6 different formats - 3 group and 3 individual.

In format 1, a student saw 4 similar, relatively common words and heard a key word spoken by the teacher. The student's task was to select the one word beginning with the same **letters** as the key word. For example, upon hearing **brooch**, the students selected from among the set, **bus, clown, brush**, and **crown**.

In format 2, a student saw a key word such as brooch, with the appropriate blend or digraph underlined, followed by four pictures such crown , clown , bus , and brush . The student's task was to select the one picture which started with the same letters that were underlined in the key word.

The third group test format, format 6 in the experiment, was the typical "fill in the blank" format used in experiment 1. For example, the student may have seen the context sentence, "I will ______ my teeth before I go to bed," and the choices **blush, rush,** and **brush.**

Formats 3 and 4, both individual formats, were tested simultaneously, with words from the two formats randomly interspersed. Format 3 words consisted of 16 familiar words, each one containing one of the target blends or digraphs. Format 4 was identical except for the fact that it consisted of 16 unfamiliar words. The student's task was to try to read each word in the 32 item list.

Format 5 consisted of 32 words (2 familiar examples of each blend or digraph) embedded in a paragraph context written especially to accommodate the 32 test words.

Procedures

Order of administration was balanced as well as possible except that the group tests were always given first. There were six different group testing orders: 126, 162, 216, 261, 612, and 621. For half the students in each group, the conbined test for formats 3 and 4 preceded the format 5 test; for the other half, format 5 preceded the combined 3-4 format. Sample items were used for all 6 formats.

Scoring for the group test formats was straightforward – the student chose the correct response or else he or she did not. For the individual formats, a target word was scored correct if the student pronounced the blend or digraph correctly but missed the rest of the word. For example, if a student said "cliff" for **clef**, the item was scored correctly.

Analyses

The overall ANOVA was omitted to allow for 5 contrasts each of which compared test 5 with one of the other 5 formats.

Difference scores (Test_x – Test 5) were computed and subjected to a multivariate analysis to test the significance of difference scores.

Finally, relationships were analyzed by computing correlation coefficients and mastery tables similar to those used in experiment 1.

Results and Discussion

The data for the individual comparisons revealed two significant pairs: Test 4 ($\overline{X} = 13.774$) differed significantly from test 5 ($\overline{X} = 14.69$), as did test 6 ($\overline{X} = 13.645$). Test 1 ($\overline{X} = 14.968$), test 2 ($\overline{X} = 14.516$) and test 3 ($\overline{X} = 14.806$) did not differ from test 5.

The multivariate test for differences between test x and test 5 was significant F (4,120) = 551, $p < .0004$, thus verifying the efficacy of the t test results.

It is of interest to note (especially in contrast to experiment 1) that the two deviant formats were unfamiliar words in isolation (Format 4) – an individual test, and the fill-in-the-blank task (Format 6) - a group test.

The intercorrelations revealed some significant relationships. Test 1 (listen and find a matching word) does not correlate significantly with any other test. Similarly test 2 (select a picture) does not seem to correlate well with any of the other tests. Even its one significant correlation with test 5 ($r = .356$) can hardly be considered important. The other 4 tests seem to be tapping some common trait. The range of correlations for those tests is from $r = .394$ to $r = .777$.

Two of the 3 highest correlations in the matrix are between test 5 and tests 4($r = .647$) and 6(.721); these two tests, even though they yield mean scores which are significantly different from test 5, seem to be assessing some traits held in common with test 5. Conversely tests 1 and 2 did not yield means different from test 5; yet they share little variance in common with test 5. Only test 3 (reading familiar words in isolation) yielded mean score results similar to test 5 while sharing a great deal of variance in common with test 5($r = .777$).

The final analysis of data compared the consistency of mastery judgments made using various tests. Comparisons were made of test 5 against each of the other tests, with a mastery criterion of 13/16 (26 of 32 on test 5). Even though the correlation between tests 1 and 5 was quite low ($r = .256$), test 1, along with test 3, produced the smallest percentage of errors of mastery judgment (test 1 > test 5 in 13% of the cases). Tests 2 and 4 tied in producing the greatest percentage of errors (29%).

Discussion

The results from experiments 1 and 2 are inconsistent; hence, any conclusions regarding the most appropriate methods for assessing word identification skills must be regarded as speculative.

Why did the traditional recognition format (#1 in each experiment) prove to be such a poor surrogate for assessing vocabulary knowledge and such a good substitute for assessing letter sound knowledge? Conversely, why did the fill-in-the blank format prove to be such a good estimator of vocabulary knowledge while it was relatively inaccurate in estimating letter-sound knowledge?

Two general possibilities exist to help explain the turnabout in the efficiency of these two formats. The first possibility has to do with the fact that the designs of the two experiments were different. In experiment 1, the recognition format was tested against a word list reading criterion while the fill-in-the-blank format was tested against a context paragraph criterion. In experiment 2, however, both were tested against a context paragraph criterion. Some data is available to evaluate this explanation. In experiment 2, one can compare the recognition format (format 1) with reading familiar words in isolation (format 3). The correlation between 1 and 3 was only .167. In percentage of mismatches, they were quite similar (each overrated mastery relative to the other in 7% of the cases). Unfortunately, the design of experiment 1 did not allow for a comparison of the recognition format with the context paragraph criterion.

One must recognize that some of the low intercorrelations between the recognition format and the other five formats in experiment 2 may be an artifact of a ceiling effect. On the average students missed about 1.5 items on format 1. However, format 3 yielded similarly high scores but correlated well with formats 4 and 5. It would be helpful to replicate experiment 2 with a sample of students less knowledgeable about consonant blends and digraphs.

The second possibility has to do with the nature of the tasks themselves. It may be that recognition formats serve as good substitutes for production tasks only when atomistic skills are being assessed. Conversely as the skills become more global (i.e., deal with larger and more meaningful units), substitute formats must themselves become more global. Such an explanation is consistent with the data in these two experiments. The recognition format served as a good estimator for the more atomistic skill of consonant blends and digraphs but not for the more global vocabulary knowledge task. The situation was reversed for the fill-in-the-blank format. The fill-in-the-blank task could not be solved (in the vocabulary study) without attention to the entire target word as well as to the context sentence. Therefore it was quite similar to the context paragraph task in terms of task requirement.

However in experiment 2, the scoring procedures allowed for correct responses in the context paragraph format when, in fact, the entire word was not read correctly. Hence the recognition format may have had the edge over the fill-in-the-blank format in terms of similarity to the paragraph reading task. We are now in the process of reanalyzing the experiment 2 data with a scoring criterion that requires correct reading of the entire word.

References

Botel, M. *Botel Reading Inventory*. Chicago: Follett, 1961.

Karlsen, B.; Madden, R.; & Gardner, E. *Stanford Diagnostic Reading Test: Level I (Form 2)*. New York: Harcourt, Brace and Word, 1966

Potter, T. & Rae, G. *Informal reading diagnosis*. Englewood Cliffs, N. J.: Prentice Hall, 1973.

Ramsey, W. & Harrod, D. Diagnostic measures of phonic analysis skills. In *Proceedings of the International reading association*. Vol. 12. Newark, Del.: IRA, 1969, 642–645.

Stevenson, J. & Johnson, D., A comparison of children's decoding and encoding responses to synthetic words. Unpublished manuscript, University of Wisconsin, 1974.

MODIFICATION OF PUPIL ATTITUDE TOWARD READING THROUGH POSITIVE REINFORCEMENT SCHEDULING

ROBERT L. AARON
University of Georgia

HAL W. SEATON
University of Georgia

ABSTRACT

A mystifying outcome of even a "well taught" reading class is that pupils who read "well" often consider reading simply as a means of attaining grades, or because the "teacher makes you."

It has been indicated that the personality of the teacher, that which she physically shows and verbally expresses, tends to be copied by students in their classroom behaviors. Also, it has been shown that reinforcing pupils tends to heighten their concept concerning that behavior, resulting in an acceleration in frequency of repetition.

This study measured whether pupil attitude toward reading could be influenced by increasing the positive reinforcement schedule of teachers. Five hundred and forty-three pupils in grades 3 through 7 in a rural, black, southeastern school district were administered the *Estes Attitude Scale* as a measure of change in reading attitude. Treatment consisted of training teachers to increase positive reinforcement activities. Baseline data were collected at 10 second intervals, on the positive and negatively reinforcing behaviors of 25 teachers. These observations on their positively and negatively reinforcing classroom behaviors were discussed with the teachers. A second rating was conducted without notice two months following the initial baseline rating. If this did not exceed the baseline rating, further counseling was implemented and specific suggestions for change were made. A final rating was carried out on all teachers.

The *Estes Attitude Scale* was administered as a final posttest of change in pupil attitude toward reading. Correlational analysis and analysis of variance were used to identify the relationships between change in teacher reinforcement schedule and pupil attitude toward reading, and to identify variations in pupils' attitudes by sex, school and teacher categories.

While there was not a significant correlation between teacher reinforcement schedule and pupil attitude about reading, some trends

did emerge when a one-way analysis of variance was applied to the mean teacher and pupil gain scores. There appeared to be a trend toward differences in both positive reinforcement schedule and attitude change favoring the third grades over the other grades, males over females, and School A over the other two schools. This suggests a significant reinforcement effort on the part of the third grade teachers at School A. This effort evidently acted to significantly influence male students more dramatically than female students in terms of attitudes about reading.

Attitude about reading is apparently far too complex a construct to be altered significantly in a more positive direction by a short-term teacher program of reinforcement scheduling. However, the trends in this study suggest that more intensive teacher effort, more attention to contriving successful pupil reading experiences, and freer delivery of "payoffs" to pupils who are reading offer some possibility of modifying pupil attitudes about reading.

RECOGNITION OF SENTENCES AS A FUNCTION OF SEMANTIC AND SYNTACTIC ORIENTING TASKS

FRED W. OHNMACHT
State University of New York at Albany

JAMES T. FLEMING
State University of New York at Albany

Studies such as those reported by Sachs (1967, 1974) have emphasized the durative characteristic of meaning over form when Ss are asked about a sentence embedded in some higher order context. A typical test involved presenting Ss with sentences for study and subsequently presenting them with a recognition test containing target and transformed distractor sentences. Ss tend to have difficulty in rejecting sentences with surface transformations, whereas, they are more able to reject sentences as not having occurred when they represent semantic modifications. The findings that surface structure information is more readily forgotten relative to semantic information are quite compatible with theories of memory which advocate discrete memory stores (cf. Wickelgren, 1972).

As a counter to the foregoing, Hyde and Jenkins (1969, 1973) have shown that free recall list effects are modified when Ss are given formal or comprehension oreinting tasks prior to study with formal orienting tasks dibilitating recall and clustering. More recently Grasesser and Mandler (1975) demonstrated that Ss did retain surface structure information when enjoined to pay attention to the wording of sentences (judge their grammaticality) whereas, such information was retained at chance levels when Ss were instructed to concentrate on semantic analysis. These findings are consistent with a levels of processing model for memory due to Craik and Lockhart (1972). They suppose that material is processed at a variety of levels and modes with the semantic being that which older subjects, at least, are usually preoccupied with. Since later access to material is dependent upon type of processing at input the fact that uninstructed Ss as in the Sachs (1967, 1974) studies retain semantic information at the expense of surface structure is consistent with their view as are the quite different results of Graesser and Mandler (1975). Put in another way, what the subject is doing at input influences recall and recognition (Jenkins, 1974).

The present study examined the effects of semantic and syntactic orienting tasks on the recognition test sentences which vary in terms of the number of propositions they contain. Bransford and Franks (1971) found that Ss tended to more frequently respond "old" to test sentences which had never occurred on the study trial as a function of an increased number of propositions or elements. We hasten to add that events described by the test sentences at varying levels of complexity all were represented in the study set.

Method

Study and recognition test sentences were identical with those used by Bransford and Franks (1971) Study II. The study set consists of 24 sentences which may be divided into sets of six which describe separate events. An example of a study set describing an event follows.

1. The hill was steep.
2. The car pulled the trailer.
3. The old car climbed the hill.
4. The car pulling the trailer climbed the hill.
5. The old car climbed the steep hill.
6. The car pulling the trailer climbed the steep hill.

Note that each ordered pair contains one, two, and three propositions. In presentation, sentences were randomly arranged and interspersed with sentences from the three other events. Study sentences were presented at a 10 second rate with a color naming task presented after each sentence. Fifteen Ss were randomly assigned to each of three conditions: a control, syntactic orientation, and semantic orientation. In the control condition Ss were presented sentences for study as previously described. The Ss in the syntactic orienting condition were asked to write down the number of nouns or verbs in the study sentence preceding the color naming task. In the semantic condition Ss were asked to write down the answer to an elliptical question such as What was? or Where? etc. after completing each sentence study – color naming task block.

Ten minutes after the study sequence Ss were all presented a common recognition task consisting of 30 sentences. Six of these sentences were non-case sentences; that is they had not occurred in the study set and involved arbitrary new information or a mixture of event facts. Two of the non-case sentences had three elements or propositions and four sentences had four elements. The other 24 sentences described aspects of the events represented in the study set at four levels of complexity representing the number of propositions contained in each sentence. None of these sentences appeared in the original list. Upon hearing each sentence Ss judged it to be old or new and rendered a confidence judgment on a five-point scale. For each type of sentence (one, two, three, and four element true event and three and four element non-case) a weighted judgment average was produced for each

S by coding old judgments as -1 and new judgments as +1 and weighting by the confidence rating and averaging over sentences of a similar type. Thus each S received six weighted false recognition scores. A constant of +5 was added to eliminate negative signs. Thus a score of 10 would indicate a S was extremely confident that a sentence had not occurred whereas, a score of 0 would mean a S was extremely confident a sentence had occured.

The foregoing resulted in a 3 X 6 (condition X sentence type) repeated measures design with conditions being a between S factor and sentence type a within S factor. The dependent variable was the weighted recognition scores which were subjected to a repeated measures analysis of variance followed by graphic plotting and appropriate multiple comparisons.

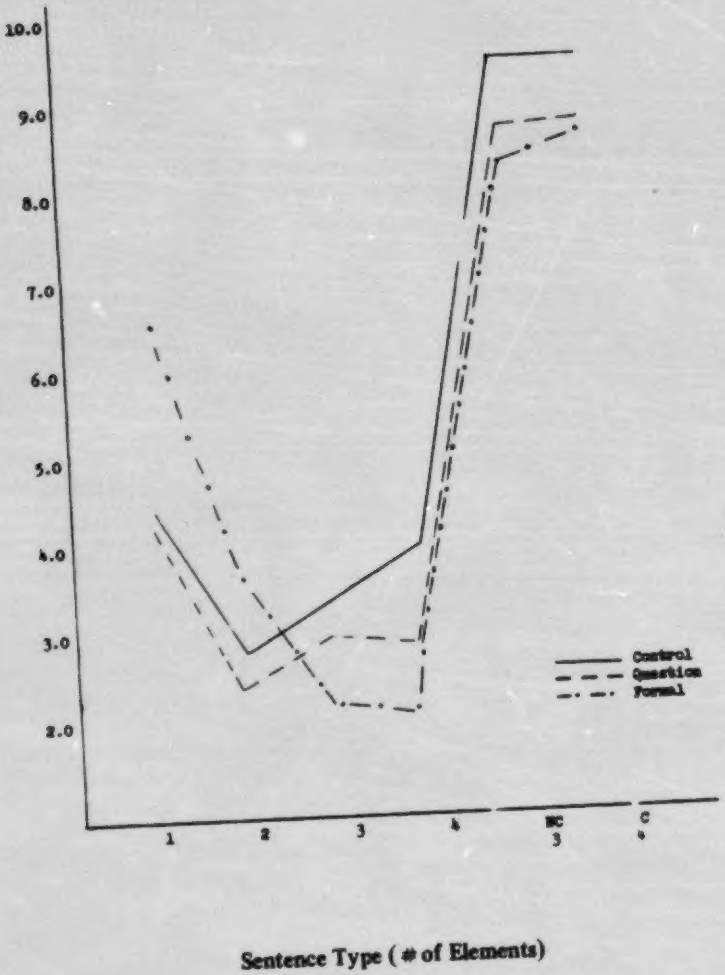
Results

Means and standard deviations for the condition X sentence data are presented in Table 1 and displayed graphically in Figure 1. Upon inspection, the means and the plot suggest a number of conslusions

Table 1
Means and Standard Deviations

	Condition Semantic		**Syntactic**		**Control**	
Sentence Type	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
1 Element	4.32	1.79	6.63	.94	4.55	2.30
2 Element	2.51	1.59	3,77	1.10	2.91	1.87
3 Element	3.15	2.31	2.38	1.82	3.44	2.25
4 Element	3.00	1.92	2.20	1.73	4.15	2.66
Non-Case 3 Element	8.87	2.56	8.50	3.03	9.60	.60
Non-Case 4 Element	8.96	2.14	8.80	1.19	9.61	.75

Figure 1
False Recognition – Confidence Means for
Condition X Sententc Type



which are formalized by subsequent analysis of variance and mean comparisons. Clearly the non-case sentences were confidently judged to be new by all groups. Of the event related sentences, there was a trend for Ss to become more confident that sentences were old as the number of elements increased although this was more pronounced for the syntactic group for three and four element event related sentences. In considering the between condition differences, it appeared that for one element event related sentences the syntactic group tended to be moderately confident that the sentences did not occur, whereas, in the control and semantic conditions Ss were marginally certain that sentences had occurred. As complexity increased for event related sentences all groups became more certain that two element sentences occurred and for three and four element sentences the syntactic group became even more certain that the sentences occurred than their semantic and control peers. All of this suggested effects due to type of sentence and an interaction between orienting task.

Indeed, these observations are confirmed by the analysis of variance for these data which is summarized in Table 2. The overall effect for orienting task is nonsignificant F(2, 42) = .81, p > .05), whereas the effect for sentence types or elements is significant F(5, 210) = 210.54, p < .001. The interaction between sentence type and orienting task is also significant, F(10, 210) = 5.38, p < .01, thus calling for a simple effects analysis to examine mean differences due to orienting task at each sentence type and of sentence type within each orienting task. These comparisons were accomplished with a series of Tukey A comparisons and are summarized in Tables 3 and 4.

Table 2
Summary of Analysis of Variance

Source	df	MS	F
Orienting Task	2	8.50	.81
Error (a)	42	10.50	
Elements	5	387.40	210.54**
Elements X Orienting Task	10	9.90	5.38**
Error (b)	210	1.84	

** p < .01

Table 3

Tukey Tests of Mean Differences for Simple Effects of Sentence Type at Each Orienting Task*

Sentence type at question	NC 4	NC 3	1 Element	3 Element	4 Element	2 Element
	8.96	8.87	4.32	3.15	3.00	2.51
Sentence type at formal	NC 4	NC 3	1 Element	2 Element	3 Element	4 Element
	8.80	8.50	6.63	3.77	2.38	2.20
Sentence type at control	NC 4	NC 3	1 Element	4 Element	3 Element	2 Element
	9.61	9.60	4.55	4.15	3.44	2.91

---

* Critical .01 value is a mean $\geq$ .90. Any two sentence type means underlined by a common line are not significantly different. Any pair of sentence type means not underlined by a common line are significantly different.

Table 3 contains the comparisons of sentence types at each orienting task. All groups present a pattern wherein non case sentences were reliably differentiated from event related sentences but not from each other. With regard to the event related sentences both the semantic and control groups were relatively homogeneous in their weighted confidence means in that for the semantic group only one element sentences were significantly different from all others whereas for the control group one element sentences were reliably different from two and three element sentences with the four element sentence somewhat anamolously placed as being different from two element sentences but with Ss being more certain that two element sentences were old than four element sentences.

Comparisons of orienting tasks at each sentence type are summarized in Table 4. For non-case sentences the syntactic group was

Table 4
Tukey Tests of Mean Differences for
Simple Effects of Orienting Task
for Each Sentence Type*

Orienting Task at 1 Element	Formal	Control	Question
	6.63	4.55	4.32
Orienting Task at 2 Element	Formal	Control	Question
	3.77	2.91	2.51
Orienting Task at 3 Element	Control	Question	Formal
	3.44	3.15	2.38
Orienting Task at 4 Element	Control	Question	Formal
	4.15	3.00	2.20
Orienting Task at NC 3	Control	Question	Formal
	9.60	8.87	8.50
Orienting Task at NC 4	Control	Question	Formal
	9.61	8.96	8.80

* Critical .01 value is a mean difference $\geq .78$. Any two orienting task means underlined by a common line are not isgnificantly different. Any pair of orienting task means not underlined by a common line are significantly different.

significantly different from the control group but not from the semantic group. The semantic and control groups were homogeneous. Note again that the means for all groups are uniformly high indicating high confidence for all groups that non-case sentences had not occurred. For event related sentences the semantic and control groups were reliably more certain than the syntactic group that one and two element sentences **had** occurred in the study list. There was no significant difference between semantic and control groups for these sentences. Further, for the one element sentences the syntactic group was midly certain that the sentences **had not** occurred. For three element sentences the picture changed rather dramatically. The syntactic group was now quite sure that sentences **had** occurred therefore, they did not differ from the semantic group but were reliably more certain than the control group. When four element sentences were considered all groups were reliably different from each other and suprisingly the syntactic group was the most certain that the sentences had occurred!

Discussion

The data summarized here are larely in accord with a depth of processing formulation for memory. Clearly, syntactic orientation questions rendered the one and two element sentences more suspect to Ss. The notion forwarded by Craik and Lockhart (1972) that Ss are normally preoccupied with processing meaning was supported by the lack of difference between control and semantic orienting groups. But why did the syntactic group become quite certain that three and four element sentences had occured? Graesser and Mandler (1972) offer the intuitively appealing notion that syntactic judgments involve semantic processing. In other words listeners will routinely store semantic information. However, when a task requires attention to surface information, Ss were able to store such information. In the present study this explains the results for one element event related sentences. As the sentences become more complex the Ss in all likelihood began to encounter an increasingly difficult discrimination problem where semantically referenced information became more salient for the Ss relative to surface information and took precedence over such information.

Several previous studies (Ohnmacht & Fleming, 1974; Ohnmacht & Ohnmacht, 1974) have reported some interesting differences between good and poor readers when memorial tasks involving word recognition and word recall are at issue. Paris and Carter (1972) have shown the applicability of sentence recognition tasks with children. Providing good and poor readers with different orienting tasks might provide some interesting insights into differential processing strategies. We would hazard the hypothesis that poor readers in control and semantic would be reliably different since they might not be preoccupied with meaning as apparently the Ss in the present study were.

References

Bransford, J.D., & Franks, J.J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, **2**, 331-350.

Craik, F.M., & Lockhart, R.S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 671-684.

Graesser II, A., & Mandler, G. Recognition memory for the meaning and surface structure of sentences. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, **1**, 238-248.

Hyde, T.S., & Jenkins, J.J. Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology* 1969, **82**, 472-481.

Hyde, T.S., & Jenkins, J.J. Recall for words as a function of semantic, graphic, and syntactic orienting tasks. *Journal of Verbal Learning and Verbal Behavior*, 1973, **13**, 471-480.

Jenkins, J.J. Remember that old theory of memory? Well,forget it! *American Psychologist*, 1974, **29**, 785, 795.

Ohnmacht, F.W., & Fleming, J.T. Developmental changes in memory attributes of good and poor readers. Paper presented at National Reading Conference, Kansas City, Mo., 1974.

Ohnmacht, F.W., & Ohnmacht, S.R. Reading ability as a mediator of effects of semantic features on short term memory for words. Paper presented to National Reading Conference, Kansas City, Mo., 1974.

Paris, S.G., & Carter, A.Y. Semantic and constructive aspects of sentence memory in children. *Developmental Psychology*, 1973, **9**, 109-113.

Sachs, J. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, 1967, **2**, 437-442.

Sachs Jr. Memory in reading and listening to discourse. *Memory and Cognition*, 1974, **2**, 95-100.

Wickelgren, W.A. Coding, retrieval, and dynamics of multitrace associative memory. In L. W. Gregg (Ed.) *Cognition in learning and memory*. New York: Wiley, 1972.

A STUDY OF THE INFLUENCE OF MINORITY GROUPS IN THE SELECTION AND DEVELOPMENT OF BASAL READING PROGRAMS: A FIVE YEAR REPLICATION

M. JEAN GREENLAW
University of Georgia

ABSTRACT

The major purpose of this study was to ascertain what, if any, changes had occurred from 1970 to 1975 in influence exerted by minority groups on schools and publishers in regard to the adoption and development of basal reading materials. A questionnaire was sent to five major publishers of basal reading programs. A similar questionnaire was sent to 75 school systems in the United States. The school systems were selected in a stratified systematic unaligned random sampling. This study is a replication of one conducted in 1970. An SPSS computer package was run, as was a McNemar test for significance of change. Because the returns in 1970 were limited, it was impossible to obtain significant statistical data.

On the basis of the results, these conclusions are suggested:

1. The increase in respondents from 1970 to 1975 indicates that there might be a growing awareness of or sensitivity to minority involvement in textbook use.
2. Development of specific criteria for textbook selection seems to be increasing.
3. Minority populations are gaining greater consideration within criteria for selection of basal readers by school systems. This is particularly true of Blacks and women.
4. Organized minority groups are exerting more influence on school systems concerning basal reading adoptions. This is particularly true of women's rights organizations, and most particularly true of the National Organization of Women (NOW), which was mentioned specifically on several questionnaires.
5. School personnel still exert the greatest influence for minority consideration in basal adoptions.
6. In both 1970 and 1975, many school systems perceived weaknesses in basal treatment of minority groups.
7. In both 1970 and 1975, few school systems made specific requests for changes in basal treatment of minorities. Most of those which did request changes noted that they spoke to local publisher's

representatives, rather than the editorial office. Notable exceptions were systems in California and Texas where they were aware that their recommendations to the state department of education were discussed with the publishers.

8. The areas of language, illustrations and content, within basal reading programs, are perceived by school systems as being weak in their treatment of minority populations.

9. Publishers perceive their basal reading materials to be inclusive of minority groups in the areas of illustrations and content, and to some extent, language.

10. There is a perceptible change from 1970 to 1975 in requests for changes in basals from school systems.

11. There is a marked change from 1970 to 1975 in requests for changes in basals from organized minority groups. All five publishers indicated pressure from women's groups.

School systems are showing increased sensitivity toward minorities, but there is still a tendency to be concerned with only the populations represented in a community.

There is still a discrepancy between the publisher's perception of fair treatment of minorities in their reading materials and that of school systems. It has been shown in this study that organized groups can bring about change in reading materials. Organized women's groups have gone straight to the source, the publisher, and have demanded change in textbooks. They have used their power to bring about this change. The same has been done by the two major state-adoption states, California and Texas. It is obvious that publishers will alter materials when sales are affected. It is up to the general school public to make their dissatisfaction known in very specific ways.

The movement reflected in this study is heartening. It indicates that we, in education, are growing toward a more representative curriculum and that we are demanding materials that reflect a more honest representation of our school population. It is hoped that if this study were replicated in another five years that the change would be even more dramatic.

ENGLISH LANGUAGE AND CONCEPT DEVELOPMENT AMONG PRIMARY LEVEL HUTTERITE CHILDREN–A CASE STUDY IN INTERCULTURAL READING EDUCATION

J. RAY STEWART
University of Lethbridge

DOROTHY M. LAMPARD
University of Lethbridge

MYRNA McGREGOR
University of Lethbridge

Of the German-speaking Anabaptist groups (including the Mennonites, the Amish and the Hutterites) who emigrated to North America from Europe in the 18th and 19th centuries the least understood is the third of these, often called the Hutterian Brethren, who first settled in South Dakota in the 1870s and then gradually moved northward into Canada. Since 1918 they have become a prominent minority group distinguished by their separatist communal life and agrarian efficiency. Since their entry to Canada about 100 colonies have been established in Alberta and substantial numbers in the adjoining provinces of Saskatchewan and Manitoba. The determination of these Brethren to isolate themselves from the political and social life surrounding them has long impeded enquiry and understanding by their neighbors. For this reason, the teaching of children in Hutterite colony schools is a unique challenge to the classroom teacher. The almost total use of an archaic Tyrol German dialect in colony life together with the lack of opportunity for the children to acquire English language skills through social experience outside the classroom produces a situation which very few teachers are prepared to meet. Compounding the difficulty already described is the fact that the young Hutterite children are learning to read and write the language of the Bible in High German at the same time that they are being introduced to English and are using Hutterite German for communication within the colony. Thus the children develop in a tri-lingual environment. Along with these circumstances the teacher finds also that Hutterite children, in conformity with their parent's value-system, do not have the career-related aspirations often found among non-Hutterites. While Hutterite parents do acknowledge the need of English literacy, they frown upon education that would encourage their children to seek personal fulfilment through the pursuit of interests or careers which might lead them away from the colony.

This factor may exert subtle influence in the motivation of the children in the English school program.

For practical reasons this report will focus on some aspects of the linguistic performance (especially phonological and semantic) of the Hutterite children.

Purpose of the Investigation

Teachers normally begin teaching assignments in colony schools with minimal knowledge of the Hutterite people or of the learning characteristics of the children. Since little has been known of the perceptual, cognitive, or linguistic development of the Hutterite children, particularly in the primary grades, it was believed by the present researchers that an investigation of the concept and language development of Hutterite children was necessary. It was expected that understandigs of great value would result, both for the preparation of teachers for this and other bilingual learning situations and for curriculum – planning specific to the Hutterite scene. At the same time, it was anticipated that the study would serve to assess the effects of the present curriculum in colony schools. Along with the purposes already mentioned the investigators hoped also they would be able to develop assessment and research techniques through the study which might be applicable to other contexts in bilinqual-bicultural education.

Procedures

Because the investigators were unable to find any research design readily adaptable to their purposes, because a major object of this exploratory study was to obtain as complete a picture as possible of the performance of Hutterite children at the grade one and two level and because the number of children varied so greatly from site to site a case study method of research was employed.

Subjects in the Study

The subject participating in the investigation were 78 grade one and grade two Hutterite children in 12 colonies of Lethbridge county, Warner county and Cardston municipal district of Southern Alberta. Since the colonies of the area belong to two slightly different groups or "leut," the Dariusleut and Lehrerleut, care was taken to have both well represented. Thirty-eight of the subjects were from Darius group colonies and 40 from Lehrerleut. The two groups vary in their first grade entrance policies with the Lehrer people providing a German kindergarten at the normal grade one entrance age of 5½ to 6½ years and the children thus entering the English school program a year later than usual. The Darius people, however, follow the customary practice of enrolling their children in the English school at the approximate age of six. The effects of these differentiated admission practices is a subject already being researched by the present investigators. In addition to the Hutterite children already described 31 pupils (16 grade one and 15 grade two) from non-Hutterite schools in the same geographic area were used as a control group.

Description of Instruments

In order to design techniques for obtaining samples of perceptual, conceptual and linguistic performance as well as measurements of school achievement an instrument had to be developed. With this intent three standardized tests of achievement: The Monroe-Sherman Arithmetic Computation Test, The Schonell Graded Word Reading Test and The Schonell Graded Word Spelling Test were administered to the grade two students. These tests all served to indicate the effects of the present curriculum on the children over a one-year period, with the arithmetic computation test especially providing a measure of learning capacity where language is not a factor. To assess both the receptive and expressive language development of the children – (phonological, syntactical and semantic) 10 categories of task were given: following oral directions given by the examiner; oral labelling of objects pointed at by the examiner (Stanford-Binet); the identification of color cards (adapted from the Peabody Language Development Kit, Level I); number card identification; word, phrase and sentence echoing (adapted from the Monroe Readiness Test); word definitions (Stanford-Binet); story repetition (adapted from the Slingerland Screening Test); story construction in response to a picture (from the Dailey Language Facility Test) and sentence completion in response to a picture presentation, such as "here is one leaf here are two ______," an item adapted from the Illinois Test of Psycholinguistic Ability. In addition to these measures the examiners administered the Peabody Picture Vocabulary Test For B to all the children as a measure of receptive vocabulary development.

All of the measures except the spelling and arithmetic tests were administered individually to the Hutterite children and control group children by four trained examiners, who also interviewed the English school teachers and in most cases the German teacher also. The former were asked to complete questionnaires about their programs. Language performance data were recorded on cassettes by the examiners. The test data were coded and computerized so that comparison could be made among the groups. As it was deemed necessary chi-square analysis was used to determine whether statistically significant differences in performance occurred between the groups.

Analysis of Data

The data on the Hutterite children were analyzed qualitatively and quantitatively using test and developmental norms where applicable. Since the number of variables assessed is far too numerous to permit analysis here only selected items are discussed. Likewise it is not possible in this presentation to compare extensively differences in the performance of children in the two types of Hutterite colony, Dariusleut and Lehrerleut.

Basic Skill Subjects

When chi-square analysis of the grade two achievement test scores was made, it was found there was no significant difference between the performance of the Hutterite children in arithmetic computation and those in the public school but the performance of the colony children was significantly lower (at the .01 level) in both spelling and reading vocabulary. An examination of the spelling errors revealed that many of the Hutterite children seemed unfamiliar with the vowel sound-symbol equivalents and in the reading test many errors occurred because the children attempted to read mainly by sight, using the configuration of the word or the initial consonant as the major clues.

The performance of the grade two Hutterite children seemed to indicate that:

1. because there was no significant difference in arithmetic performance between the Hutterite groups and the public school group, the Hutterite childreh had basic learning competence and could perform as well as non-Hutterites.
2. because there were significant differences in spelling and reading scores between the Hutterite and public school groups, there must be factors, other than intellectual ability, which are interfering with the performance of Hutterite children in these areas.

Language Measures

Phonology. In this area of observation an attempt was made to determine the progress the Hutterite children had made in the acquisition of the English phonemes, the smallest sound segments distinguishing one word from another. Students were asked to "echo," that is, repeat after the examiner, words, phrases and sentences containing the English phonemes being examined. The responses were recorded and later analyzed. If a student accurately echoed the phonemes of the examiner, this was taken as evidence that he could both distinguish the phoneme in speech and produce it, although he might not always, in every verbal exchange, produce the phoneme correctly. Because the Hutterite students are native speakers of a dialect of German, it was thought that there might be some interference between the phonemes of the two languages. The phonemes which produced significant differences between Hutterite and public school students, as well as the phoneme commonly substituted by the Hutterites were:

Phoneme	Substitution
/ ð / as in then	/d/
/ θ / as in thread	/t/
final /r/ as in better	omitted or / ə/ as better
final /t/ as in quite	left out or /d/
/ɔ/ as in bought	/o/ as in note
/a/ as in father	/o/ as in note
/æ/ as in bat	/ ɛ / as in bet or /e/ as in bait

In addition, the performance showed that there is a significant difference in the ability of Hutterite and non-Hutterite students to echo such words as "transcontinental" which are probably not within the vocabulary of either group. The problem with the multi-syllabic words may reflect an overburdening of memory which at this stage must be used to recall specific individual phonemes and cannot be used to remember sequence of phonemes as well. The public school children, who are more familiar with the English phonemes, do not require as much "memory space" for these and can concentrate on sequence. Nevertheless it should be noted that the performance of the public school students was also below what one would expect from a person who was familiar with the word "transcontinental" and who could retrieve the word immediately upon hearing the stimulus.

Overall, the Hutterite students' performance seemed to indicate that they had the underlying competence to produce most phonemes of the English language. Some phonemes, such as /ð/, /θ/ and final /r/ do not occur in their dialect and, therefore, appear to be more difficult for them to acquire. There are also differences in the vowels, which are reflected in their difficulty with / /, /æ/ and /a/.

Semantics. The semantic system is here understood to mean simply the attaching of meaning to words and to strings of words. The development of vocabulary demands a growth in variety of experiences with an object so that the word begins to have a distinct meaning apart from the word for similar objects, and yet encompassing all the meanings around the word. For instance, water may initially be only something to drink or in which to bathe, but the child will experience water in puddles, lakes and streams and will see that fish live in water and certain birds swim in it. He will see that it can be dirty or clear, but it remains water. Milk, juice, and gasoline, although they have some commanlities with water, are not water. Thus a child learns not simply to attach a verbal label to an object by imitation and repetition, but to build his semantic knowledge from his experiences.

In the following section, performance in receptive vocabulary, in response to commands and in labelling tasks will be analyzed.

1. Receptive Vocabulary. The Peabody Picture Vocabulary Test, which requires the student to point to 1 of 4 pictures which represents the word spoken by the examiner, is a measure of receptive vocabulary. An examination of the scores on this test at the first and second grade level indicates that no Hutterite student at the first or second grade level has a receptive vocabulary, as determined by the norms of The Peabody Picture Vocabulary Test, greater than that of a child of 5 years 11 months. Furthermore, at the grade one level, over two-thirds of the Hutterite children had a receptive vocabulary at or below the third year level. An examination of individual responses showed that only one child, a grade one child from the Lehrer group, appeared to have no receptive vocabulary whatsoever. All other children could correctly respond to the initial words such as table, bus, horse and dog. The words at the most elementary level which were most difficult for the Hutterite children, were the present participles (e.g., climbing, pulling, hitting, ringing and baking). There was some improvement in recognition of these participles at the grade two level, but some, especially "pulling," continued to be poorly understood. It is interesting to note that "ringing," a hand bell, a normal colony event, was recognized more frequently than the other participles.

2. Response to Commands and Performance in Labelling Tasks. At the grade one level, over half the Hutterite children could not respond correctly to commands given (performance significantly different from

that of the public school students). However, by grade two, there was no significant difference in response. Labelling (the ability to attach names to objects, numerals or colors) is seen to be one of the first indications of semantic growth which can be assessed. In this area at the grade one level significant difference (at the .01 level) was found between the Hutterite and non-Hutterite performance. At the grade two level, although there was still a significant difference, the discrepancy in scores was much less. The incorrect responses are interesting and indicate that the children have not sufficiently differentiated their knowledge of the environment, at least in terms of the English language. For instance, the doll received such labels as baby, Betty, girl, and man. The dog was labelled Flip, bird, and pony. These responses seem to indicate that the child giving them is beginning to develop an idea of the object, but has not had enough experience with it and around it to completely understand it.

To illustrate differences of performance in number card labelling the numeral seven is here selected. It was noted that only one-third of the Darius group could label the numeral seven. Less than one-third of the remaining Darius students recognized the numeral when presented with the name and several examples from which to choose. An interesting contrast is shown in response to the request, "Give me seven blocks." All but two Darius students responded correctly, whereas, only five had correctly labelled the numeral. Conversely, seven Lehrer students could not accurately count seven blocks, whereas, only two could not label the numeral seven. These differences could be due to (a) the preschool training of Lehrer children which might include numeral recognition and/or (b) a more practical, concrete, commercial outlook among the Darius group (e.g., some Darius children visit the city more often with their parents and come into contact with commercial exchanges). By grade two there were no significant differences between Hutterite and non-Hutterite groups on either of these tasks, because all students had mastered them.

Summary and Implications

From the case study consideration of the Hutterite students, we derive a composite picture of a group of children who can learn as shown by their performance in arithmetic, and who do have a beginning competence in the English language. In spite of proven capacity for learning, this group of children did not perform well in the aspects of language measured. It was noted that:

1. Hutterite children do not have serious problems with English phonology but they do have difficulty with polysyllabic words and certain phonemes which do not occur in their dialect.

2. Hutterite children are at a preschool level of English vocabulary development and appear to lack sufficient concrete experiences of a meaningful kind to adequately differentiate their vocabulary (e.g., to differentiate between doll, baby, and boy).

In spite of these differences Hutterite children are exposed to the same curriculum and use the same materials as the public school children. It was noted that, in the arithmetic series commonly used in the colony schools visited, concepts are carefully developed by going from the concrete to the symbolic experience. This type of teaching appears to be suited to the level of development of the Hutterite children and they do well with it. However, in the language arts area such a progression does not occur and children are plunged into the program without the development of the necessary concrete and readiness activities.

The disadvantage resultant from the common ignorance of Hutterite culture among beginning teachers on the colonies has been well recognized in Mann's (1974) research. It is evident from the present study that recognition of the differences in the perceptual, linguistic, and cognitive development of Hutterite children is a first step to necessary curriculum modification and improvement in the performance of primary grade Hutterite children in their English school program. At the same time it may be inferred that similar analyses of the linguistic development of children in other bicultural - bilingual situations would be very profitable for the improvement of instruction.

References

Mann, G.A. *Functional Autonomy among English School Teachers in the Hutterite Colonies of Southern Alberta: A Study of Social Control.* Unpublished doctoral dissertation, University of Colorado, 1974.

ACQUISITION OF CONCRETE AND ABSTRACT WORDS IN A BEGINNING READING SAMPLE

GEORGE MCNINCH
University of Southern Mississippi

MARK G. RICHMOND
University of Southern Mississippi

SONDRA KUTZMAN
University of Southern Mississippi

ABSTRACT

The purpose of the study was to investigate the differences existing between the rate of acquisition of lexical versus structural words in a prereading sample categorized under two levels of reading readiness functioning. It was hypothesized that beginning first graders classified as either high readiness or average readiness would learn concrete words more quickly in a trials-to-criterion task than they would learn abstract words. Concrete words were defined as nouns that had tangible referents or representations and were in the receptive vocabulary of the typical beginning first grader. Abstract words, for this study, were defined as either structural words (connectives) or lexical words without tangible referents or representations.

To evaluate this word learning hypothesis, eight four-letter words (four concrete: calf, park, four, dust; four abstract: with, good, been, luck) were selected from the Dale List of 769 Easy Words according to the following factors: (a) word length, (b) configurational elements; (c) different initial and final letters; (d) freedom from obvious discrimination confusion; (e) and words not used in the beginning basal materials in the cooperating system. To produce the learning condition, the examiner flashed individual words in a predetermined random order by sets (concrete or abstract). The children were first screened to determine if they knew the words (trial 1). If the child did not know any of the words he was then individually "taught" the words by a prompting technique until all words were known as evidenced by a trial of four correctly pronounced words. Trials were discontinued if more than 30 repetitions were needed for learning. Each pupil had two scores: an abstract score and a concrete acquisition score. Task words by sets were learned at two day intervals with random assignment to the list to be learned first.

Treatment means for achievement groups indicated no difference between above average and average readiness groups ($\bar{X}$ above average =

10.94, $\bar{X}$ average = 10.38), indicating that each group required about the same number of learning trials. While the numeric difference was greater between mean scores of concrete vs. abstract words, ($\bar{X}$ concrete = 9.45, $\bar{X}$ abstract = 11.77) there was still no significant difference relative to word category.

From the data available, it does not appear that beginning first-grade pupils learned concrete words more readily than they learned abstract words in the trials-to-criterion learning tasks. Contrary to popular beliefs in teaching methodology, the findings of this data base then strongly suggests that concreteness and abstractness factors may not be as significant as previously believed in word learning.

A STUDY ON THE CHILD'S CONCEPTION OF A SPOKEN "WORD"

JERRY L. JOHNS
Norther Illinois University

ABSTRACT

The purpose of the study was to partially replicate and to expand a recent investigation on the child's conception of a spoken word. These stimuli consisted of five examples in each of the following eight different classes of auditory stimuli: non-verbal abstract; non-verbal real-life; isolated phonemes; isolated syllables; short words; long words; phrases; and sentences. The 40 stimuli were recorded by a female adult on audio-tape in four different random orders. Each subject was randomly assigned to one of the orders.

The sample consisted of 120 children selected so that there would be 20 boys and 20 girls in each of the following age ranges: 5.6 to 6.5 years; 6.6 to 8.0 years; and 8.1 to 9.5 years. All of the children were white and predominately middleclass to upper middleclass.

An analysis of the results indicated that by the time the children reached the age range of 8.1 to 9.5 their concept of a "word" was generally good except for confusing isolated phonemes and syllables as a "word." There was, nevertheless, a considerable range of individual differences when the data were analyzed by the number of children who consistently knew whether a particular class of stimuli was a "word."

By the time the children reached second or third grade (8.1–9.5) at least 80% were able to distinguish non-verbal abstract; non-verbal real life; phrase; and sentence stimuli from words. Confusions still existed among phonemes, syllables, and long words as "words."

The findings of the study generally support the results from related studies (i.e., young children do not have an adequate concept of what constitutes a spoken word). This confusion exists whether the children are British, Canadian, or American. This state of "cognitive confusion" may partially explain why some children have difficulty making sense out of instruction aimed at helping them to process print.

THE RECOGNITION OF WORD BOUNDARIES BY ADULT SECOND-LANGUAGE SUBJECTS

DAVID C. JOHNSON
U. S. Air Force

ALBERT J. KINGSTON
University of Georgia

Language instructors commonly assume that students studying English as a second language automatically recognize the words which serve as units. Recent studies by Karpova (1966), McNinch (1970), Meltzer and Herse (1969), Kingston, Weaver, and Figa (1971) determined that children sometimes had difficulty in identifying printed words in narrative prose. Kingston, Weaver, and Figa (1971) also reported that first-grade children had difficulty in distinguishing words, utterances and sounds. These studies have caused reading and language teachers to reconsider the needs and previous learning of beginning readers.

Adults who learn English as a second language often encounter difficulties in achieving competency. Traditionally teachers have focused on the structure, pronunciation and vocabulary problems encountered by the adult learner. Often the foreign students' language skills have been compared to those of native English speaking children in evaluating their learning progress. Johnson (1974), for example, found that foreign officers performed at a lower level than native speakers when evaluated by the cloze procedure. His subjects were comparable to elementary school children in English achievment.

The study was developed to provide an even more base measure of language ability, namely to assess the ability of non-native speakers of English to identify word boundaries in three different modes of presentation. The subjects were 83 male foreign Air Force officers from 35 different countries, who had been assigned to the United States for special training.

Procedures

All officers had successfully scored 65 or better on the English Competency Level Test (ECL) developed to assess English Language proficiency. The men ranged from 28–43 years of age. Five broad language groups were represented by the 83 officers. They were Hamito-Semitic, Indo-European, Sino-Tibetan, Malaysia-Polynesian, and Romance (Yuen Ren Chao, 1968). Three written scripts were represented: Sanskrit, Arabic, and Roman. All groups had approximately the same competency in English. Specially prepared materials with a readability at the 6th grade level were employed in the three modes of presentation.

The first presentation was similar to the Cowgill (1963) and Miltzer and Herse (1969) studies, but modified so that all capitalization and punctuation cues were removed. In this mode of presentation printer's justifications were left between all words. Subjects were told to draw a slash mark or line between each word of the text and were given five minutes of instruction to answer that they understood the task. Fifteen minutes was allowed for the assigned task. There were 108 correct responses as represented by word decisions.

The second mode of presentative was similar to that employed by Klein and Klein (1972). Material was presented in lower case letters with all punctuation deleted, a space was presented between each letter and no additional spaces made between words. The testing conditions and practice sessions were identical to those in the first presentation.

The final mode of presentation was similar to previous experiments by Karpova (1966) and McNinch (1970). In this study the materials were pre-recorded on tape and transmitted aurally to the subjects. The officers were asked to indicate the number of words in each sentence by writing the number on a specially prepared answer sheet. The subjects were given practice in checking or tallying each word heard and then adding the total number of words in a particular sentence. Sentence length varied from five to 18 words in length.

Results

Scoring for each of the modes of presentation was in number of errors and each of the three tasks had the same potential number of errors.

Table 1

Mean, Variance, and Standard Deviation for Conditions a, b, and c.

Condition	Mean	Variance	Standard Deviation
Printer Space (A)	2.8 errors	7.26	2.69
Embedded Word (B)	6.5 errors	13.29	3.65
Aural (C)	10.1 errors	9.70	3.11

As shown by Table 1 the aura mode of presentation proved most difficult for the subjects. The subjects averaged 10.1 errors when asked to count words in narrative material presented aurally as compared with materials presented in written form. The subjects also found it more difficult to identify words which were deliberately embedded thru the spacing of each letter, although this task was evidently less difficult trace identifying words in aura presentation. These differences were significant beyond the .01 level, $F(2,21) = 35$, $p < .01$. A t test performed on the three conditions (A vs. B; A vs. C; B vs. C) also indicated significant differences between each comparison.

These results demonstrate that the English second-language adults had little difficulty in recognizing words when they were presented in traditional print but had problems with the aural presentation. These results differ from those obtained by Meltzer and Herse (1969) with children. Presumably, the maturity of the adults or previous training in printed English or their native language enabled these subjects to utilize the printer's spacing as a cue in identifying words.

References

Chao, Y. R. *Language and Symbolic Systems*. Cambridge University Press, 1968.

Cowgill, W. A. Search for universals in Indo-European diachronic morphology. In J. H. Greenberg (Ed.), *Universals of Language*. (2nd ed.) Cambridge: The MIT Press, 1963.

Johnson, D. C. The evaluation of comprehension when English is a second language, by use of the cloze procedure. Paper presented at The National Reading Conference, Dec. 1974.

Karpova, S. N. Osoznanie slovesnogo sostava rechi rebenkom doshkol'nogo vozrasta (The preschooler's realization of the lexical structure of speech). Voprosy Psikhol. 1955, 4, 43–55. Abstracted by D. I. Slobin, In F. Smith and G. Mill (Eds.) *Genesis of Language*. Cambridge: The MIT Press, 1966, pp. 370–371.

Klein, H.A., & Klein, G.A. Studying the use of context for word identification decisions. Paper presented at the National Reading Conference, New Orleans, December, 1972.

Kingston, A. J., Weaver, W. W., & Figa, L. E. Experiments in children's perceptions of words and word boundaries. *Twenty First Yearbook of the National Reading Conference*, 1971, 91–99.

McNinch, G. H. W. The relationship between selected perceptual factors and measured first grade reading achievement. (Doctoral dissertation, the University of Georgia) Ann Arbor, Michigan. University Microfilms, order No. 71–3756. *Dissertation Abstracts*, Feb. 1971, 31, 39–65A.

Meltzer, N. S., & Herse, R. The boundaries of written words as seen by first graders. *Journal of Reading Behavior*, 1969, 1, 3–13.

LOGIC AND MEANING OF LANGUAGE IN WRITTEN COMMUNICATION

JAMES A DINNAN
University of Georgia

ROBERT A. LODGE
Kings College

Written communication is an observable, and possibly a permanent, production of a human who attempts to share with another human his or her observations of the nature of all beings. During communication, the author gives observations of the matter found in the universe. This matter is identifiable through its qualities and possesses both weight and volume. Thus a man or woman possesses the quality of humanness. They belong to a class that has been identified by a given language group using the verbal combination of agreed upon sounds or symbols to produce the spoken or written forms (man-woman). Through these agreed upon symbols the author further wishes to share his observations about man or woman in relation to all other beings. Thus the "R" words are used to show specific relationships between various matter in space. Further, the author uses a time reference, which is a measurement tool for space. Therefore, a composite statement of written communication is that it is an author's shared experience which assists other individuals in perceiving relationships of matter in time and space so that they too might be aware of the order and structure of the universe. The written symbols and their combinations consist of letters, marks, words, sentences, paragraphs, and themes.

Letters are man's feeble attempt to represent the various combinations of mouth and air movement in written form. They are supposed to be visual symbols of internal and external manipulations of a given language code. Often, they are just arbitrary symbols which a group agrees upon as the representations of an idea about matter or a given relationship (e.g., eye).

Marks have been used to signal stress, pause, stopping, questions, and the like and are symbols with meanings just as all words have specific meanings in a communication.

Words are ideas used in external communication to represent either matter or a given relationship in time and space. The nucleus of language, 110 words, accounts for 50% of language usage, (Coleman, 1968). Seventy-seven percent of the core of our language consists of references to space and time. The other 23% consist of words referring to matter (e.g., I, me, us, man, thing). Words or ideas can only be known through a contrast and contrasts are available for every word in the lexicon. Moreover, they are within one standard deviation from the original concept (e.g., **was** #9 – **is** #13, **only** #79 – **any** #80, **go** #62–**come** #98).

The essence of communciation lies in the sentence. Here we find the opportunity to view the relations set up by the author as he perceives matter and relates it in time and space.

1. The man went to the bank.

In this sentence, the matter references ("M" words) are man and bank. The remaining words show the relationship between the two matters, the (one), went (time), to (direction in space), the (one). The totality of this sentence is the relations of matter to matter in time and space. When further relations are added, the clarity of the communication improves.

2. The old man went to the bank at Fifth and Main.

Even further clarity of the precise matter might be given as in the following sentence:

3. The old man, Mr. Jones, went, at three o'clock, to the First National Bank on the corner of Fifth and Main streets.

There are still only two pieces of matter referred to in this communication in time and space. Thus an essential characteristic of the sentence is the use of relationships. This relationship can only take place when a sufficient number of "M" or "R" words are used.

The paragraph develops and clarifies by making relationships concerning the prime matter being talked about in it (Prime M or Prime R). Thus all other sentences show further relationships to either Prime M or Prime R and in this way contribute a clearer relationship to the main idea being presented.

In dealing with a composition or theme, we first find the Prime M or R which is being considered and then examine each paragraph to determine how it is showing relationships to the Prime matter or relationship of the theme. In many instances the Prime M or R is revealed in the title.

In summary, written communication uses words, either matter references or relationship references, in sentences which present a basic awareness of the matter or the acts of matter in the universe. The sender and receiver of the communication, based upon common language and experiences, share in a uniquely human act: a written communication, namely **reading**. If communication is to take place, both the sender and the receiver must be made aware of the logical development of thought and its use of precise references to time and space. Further, when references are used, by an author, they represent one half of a common contrast set and both the sender and receiver must use the same "set" to communicate basic relations. Meaning can only be achieved through a common base reference to all words found in the language not just the ones used to identify matter. The relation words have meaning which is paramount to communication. Without relation words we would be like parrots who connect strings of words. Relations in time and space of matter are known only through those agreed upon words. They must be taught with specific meaning in contrast, never as isolated items which is a current practice in most of our schools.

The first step is to review the core of our language, decide whether the word is referring to matter or a given relationship, locate its contrast (some have several), demonstrate its meaning, and present them as contrast pairs to be used as automatic responses in communication. "Automatic contrasts" are necessary because in communication the sender assumes that the receiver will use the same basic "set" as he or she does. If the receiver uses another idea or response, the communication will break down because the common denominator is missing. Both have completed an intellectual process and both have produced a product, but **no communication** will take place unless a common set is used. Automatic responses should be checked in the oral language readiness stage preceding the reading act. Contrasts should automatically be given when a single stimulus is supplied (e.g., in/**out**).

Since the core of the language is so highly redundant it must be understood as a common "set" of meaningful references not just "connector words" or "articles," or "conjunctions," but as a reference to specific time or positions in space which will show the relativity of one matter to another matter. These words are not only the core of the language but are the core of meaning. They make communication possible.

References

Coleman, E. B. Words Rank Order as to Frequency, Thorndike and Lorge. Unpublished paper presented at the National Reading Conference, December 1968.

AN ORAL LANGUAGE INVENTORY FOR REMEDIAL READERS: GRADES 1–5

JOYCE M. HAMRICK
Troy State University Dothan/Fort Rucker

ABSTRACT

There is a need to investigate the effects of and possible need for appropriate language patterns to facilitate the reading process. The present investigation was done through the use of an oral language inventory administered to a select group of third grade students.

The questions posed for study were concerned with the use to which the data collected could be put:

1. Can the use of an oral language inventory identify and isolate problem areas in oral language production;

2. Is it possible, with results from the inventory, to remediate areas of oral language weakness?

3. Will identification of and appropriate instruction in oral language production produce positive gains in standardized reading and achievement test results?

Results of a longitudinal study of this nature cannot be immediately known. Preliminary data suggested the following implications:

1. Paradigmatic responses are developmental.
2. Oral language production is sequential.
3. Convergent and divergent responses are influenced by experience.
4. Grade level placement does not ensure appropriate oral language production.
5. Oral language responses can inappropriately cross oral language production levels.
6. Incorrect oral responses do not necessarily produce inappropriate standardized test scores.

The results of this Inventory will enable the teacher to better evaluate a child's understanding of the phonetic system and comprehension skills. The inability to visually discriminate between letters can also be observed.

Results obtained on the Oral Language Inventory post-test can supply the teacher with information regarding further language instruction needed and specific areas of weakness for possible referral to appropriate sources.

Another benefit of the inventory lies in the information to be gained for use in incorporating vocabulary development with oral language production.

The ultimate aim of the Oral Language Inventory is to provide the

classroom and special reading teacher with an instrument for more accurately evaluating oral language production in remedial readers. The purpose of such evaluation is to aid the teacher in providing effective instruction in reading skills in the hope that she can produce more skillful readers.

DEVELOPING A SKILLS – BASED APPROACH TO READING COMPREHENSION*†

WAYNE OTTO
University of Wisconsin-Madison

THE DESIGN

The Wisconsin Design is not meant to be a "complete reading program." It offers the skills and objectives that can serve as a framework for such a program but it does not prescribe the program format, content or procedures. Nor is it meant to be a comprehensive program in the classic tradition of the basal readers. Again, the Design offers the skills and objectives for organizing such a program, but it does not prescribe content. The Design focuses on the skills that are judged – at this point in time – to be essential to success in reading and it provides the backup found to be essential to teach the skills. But responsibility for developing a total reading program and for seeing to it that skills are applied to content rest squarely with school administrators and teachers who use the program.

THE COMPREHENSION ELEMENT

Even a cursory search of the literature will yield dozens of lists of comprehension skills. Every basal reader and every curriculum guide will yield still another. But examine most of them closely and you have reason to hold skill lists in low esteem. The skills listed tend to be so general or so specific that they can serve no function in teaching or learning, more often than not the essential **literal** aspects of comprehension are neglected in favor of the more esoteric interpretive and creative aspects, and there seems to be no concern whatever for whether the skills are in fact teachable. If I thought that was the best we could do, I would cast the first stone to put skill lists and skill listers out of their misery.

Once we concern ourselves with sensibly chosen skills – skills that are both important and teachable – I think there will be less tendency to view them or to teach them in isolation. I see skills and teaching of skills as the **substance** and the **means** for sharpening the process of developing reading comprehension. Skills must be identified, taught, and applied in context. In perspective skills can be the vehicle for moving children to independence in their reading comprehension. Without the explicit focus of skills, I'm afraid we abandon too many learners while they are still dependent on the crutches of "classroom questions," "directed reading activities" and other alternatives to explicit skill development. The alternatives, too, need perspective.

---

*This paper represents only an abstract of the original paper. The original may be obtained from the author.

† All components of the Design are now published by National Computer Systems/Interpretive Scoring Systems, 4401 West 76th Street, Minneapolis, Minnesota.

Developmental Edition

The Comprehension element of the Design is now at a stage of development where it is available for implementation in qualified schools but still undergoing revision. While one might argue that a product ought not to be released for implementation before it has been fully and finally developed, practical considerations make a Developmental Edition desirable. Very bluntly put, funding sources, potential users and publishers want to see a product in the field at the earliest possible date. This is particularly true in the area of reading comprehension because the options that are available are neither numerous nor attractive. The developers on the other hand, would prefer to pilot, revise and perfect until they are satisfied that the product is as good as they can get it. The Developmental Edition, then, is a compromise that takes the product at a point where it is conceptually respectable and implementable and makes it available to users who know what they are getting into. The understanding is that development and refinement will continue – making use of data obtained from try outs of the Developmental Edition.

Specific Skills and Objectives

Six strands of Comprehension skills were identified in the Developmental Edition. The strands and the levels at which skills/objectives are identified are given in the schema that follows. A rationale for each strand is given in the *Teachers Planning Guide-Comprehension* for the Developmental Edition (Chester, R.D. et al., 1974). The actual skills and objectives are given in the guide and in the *Rationale and Guidelines* (Otto & Askov, 1974).

	Level						
Strand	A	B	C	D	E	F	G
Main Idea	X	X	X	X	X	X	–
Sequence	X	X	X	X	X	X	X
Reasoning	X	X	X	X	X	X	X
Detail	–	X	X	X	X	X	–
Context Clues	–	–	–	X	X	–	X
Affixes	–	–	–	–	X	X	–

The skills and objectives for the Developmental Edition are very much a product of the pragmatic-empirical approach. After a general review of research and practice related to the teaching of comprehension skills, we identified the six strands that are outlined in the schema above. As we saw it then, these strands reflected the major subareas or clusters of skills presented in the literature. In developing each strand, we attempted to identify a sequence of skills that would be compatible with the grade/skill level model we evolved for the Design. For some strands we were able to identify and state seven discrete skills/objectives and to arrange them – at least to our satisfaction – in an ascending order of difficulty. For other strands, we did not feel we had seven discrete skills/objectives. Then we simply inserted what we found into the leveling sequence at the point most compatible with the development in the other strands. Then we tried out what we had and made changes as they seemed appropriate.

ASSIGNING SKILLS TO LEVELS

As I've pointed out, skills in the Design are clustered at seven levels which are roughly equivalent to the seven levels – kindergarten through grade six – of the elementary school experience. The intent is not to lock levels to grades; but we have found that it is essential to have some well established referent to guide the placement and development of skills and objectives. The grade level referent is the best we have.

We have no serious difficulty in assigning skills/objectives to levels on the basis of logical and/or content analysis. The levels we assign get the consensual support of consultants, including classroom teachers. But when we get data from the criterion referenced tests designed to assess mastery of the objectives, sometimes we get inversions or, occasionally, breakdowns in mastery performance. The pragmatic does not always square with the empirical!

DEVELOPMENT WITHIN STRANDS

In view of concerns I have already expressed and information we have gained, we are inclined to look again at the specific skills/objectives and their arrangement within strands. Perhaps the best way for me to explain how we have proceeded is to give an example of what we are doing. The example that follows is adapted from an example I developed for a paper explaining the pragmatic-empirical approach to research in reading (Otto, 1975).

> In the fall of 1974, the Main Idea strand amounted to six objectives to be taught and mastered in the elementary school. The most basic objective had to do with the identification of the topic of a picture and the most advanced had to do with the identification of the main idea

of two paragraphs, which might or might not include explicitly stated topic sentences. The objectives were based heavily on research done earlier by the Laboratory for Research in Basic Skills at the University of Wisconsin-Madison. Data from limited tryouts indicated that children could attain the objectives and that the objectives were acceptable to teachers. To be brief, the strand looked respectable and teachers could use it.

But we wondered a number of things.

1. To what extent should the strand be reflecting the great mass of research findings that are accumulating with regard to advance organizers, cognitive organizers, paragraph organizations? Certainly there was no lack of existing research. If anything there was too much, with diverse approaches and orientations and sometimes conflicting results. Lots of promise, but perhaps as many questions as answers.

2. Is it easier to identify the topic (one or two words) of a selection than a more explicit title? Practice suggests that it is, but no data were available. Is it harder still to identify a main idea statement? What if a topic sentence is provided? Again, practice seemed to be unsupported by data and practitioners shared our concern.

3. Are children who can identify a main idea statement also able to **state** a main idea? Common sense could take us either way in answering that one. Again, no real data available. And, again, practitioners wondered, too.

We had to make some decisions. A change in objectives would mean changes in criterion referenced tests and in supporting materials for teaching and implementing the strand. All would have to be accomplished within the time/resource allocation. We decided that our reservations were serious enough to warrant changes, so this is what we have done or are now in the process of doing.

1. We reviewed the existing research relating to paragraph organization and the use of various types of organizers. On the basis of all that we decided that the objectives in the strand probably should prescribe – or at least describe – the organization of test passages as well as call for identification of a topic. We wrote seven objectives – one additional – for the strand to reflect organization of passage as well as expected outcomes. To the practitioner, this means that while the strand is being taught over the elementary school experience, explicit attention will be

paid to what appear now to be important aspects of organization. To the researcher, this means that some rather tentative suggestions from research are being tried out and need, in turn, to be examined as they are presently operationalized. (Either that, of course, or that the suggestions be further developed – which could mean they would become contraindicative of present suggestions. This would be the traditional approach and would keep outcomes out of any wide scale application indefinitely.) To the developer this means putting the whole strand together again with a new thrust.

2. We have done a study to find out whether topic, title and main idea statement do indeed differ in difficulty. They do; and this information has been looped directly back into the development of the strand in terms of how the objectives are stated and sequenced.

3. We have done a study to find out whether children who can identify a main idea statement when it is given can state a main idea on their own. We've gotten far enough with that one to conclude that they can not – or at least do not come up with statements that are definitive enough to be considered adequate. So we are adding a top level objective to the strand. The child generates a main idea for a passage that may or may not have an organizer and has both relevant and irrelevant details. (**Happy Note**: We have done this despite the fact that this virtually precludes a paper-and-pencil criterion referenced test.)

As we put the strand back together, we operationalize each objective by (a) devising an appropriate criterion referenced test, and (b) deciding what materials and procedures we can identify and suggest to help children reach the objective. In that process we work with teachers and children routinely. We send drafts of the objectives out for review by educators who are knowledgeable in the area. We try to use common sense and we check it out when we see a need. We try to keep it all palatable to practitioners and we check that out constantly. And we go out to the schools to see whether it works. Next we'll try to find out whether it does any good. We employ the pragmatic-empirical approach.

We have followed a similar procedure in looking at each of the strands included in the Developmental Edition. As a result of that effort, we have decided to do a major revamping of the substrands of the Comprehensive element. The strands are now being developed as shown in the schema that follows.

COMPREHENSION STRANDS

Word Meaning	Sentence Meaning		Passage Meaning	*Sequence
*Word parts (affixes) *Context clues	*Detail Paraphrase	Sentence Structure	*Central Thought (main idea) *Relationships/ Conclusions	

*Major strands in the Developmental Edition.

The most apparent change is that we are now focusing explicitly on Word Meaning, Sentence Meaning and Passage Meaning as aspects of literal comprehension. This is clearly in line with what the psycholinguists are saying; but, more important, we have found it to be conducive to the coherent development of the skills within a strand. Consider, for example, the Sentence Meaning strand. With the focus now clearly on the meaning conveyed by sentences, we are developing the strand to deal not only with the **detail** aspect but also with sentence structure and paraphrase. This, we feel, moves us to where we are in fact providing teachers and learners not only with essential skills/objectives but also with a better vehicle for pursuing them. The papers that follow in this symposium deal specifically with the development of the revised strands.

There are other changes. We have combined the old Affixes and Content Clues strands into a single Word Meaning strand. The sense of that should be obvious. We have renamed the old Main Ideas strand Central Thought to better reflect what really goes on when the skills are applied. We have given the old Reasoning strand a major overhaul in terms of both terminology and conceptualization and renamed it, tentatively, Relationships/Conclusions. Frankly, old Reasoning presented our biggest problems in terms of credibility, teachability and assessibility. The new treatment moves us forward substantially in solving those problems. We have combined both Central Thought and Relationships/Conclusions in a single Passage Meaning strand. The name was not change for the sequence strand, but the objectives have been restated and reorganized to deal more effectively with instructional concerns.

The changes are both apparent and real. We have not thrown out anything, because we learned from our experiences. At the same time, we have not preserved anything for its own sake, because we have also learned enough to move forward.

OVERLAP OF SKILLS/OBJECTIVES

The fact of overlap among skills and objectives in the different Comprehension strands is real and it has been recognized from the start.

One cannot say with much conviction that, for example, the Context Clue skills are wholly discrete from the skills in the Affixes strand. Nor can we claim that the Main Idea objectives are unrelated to the Detail objectives. And, when we examine the intercorrelations of scores on the criterion referenced tests for skills in different strands, we find the coefficients are both positive and substantial. All of this suggests that what we may be doing is simply tapping some global, underlying "skill" with differnt measures; so each time we sample we come up with something that is quite similar to what we got before in terms of performance. The cynics—or you might call them the purists if your views are similar—have said this means there are no subskills for Comprehension. They suggest that we either (a) discover skills that are discrete and unrelated, or (b) abandon and skill-centered approach and simply teach "comprehension."

The problem that confronts us is not whether there is overlap among the skills we are identifying. There is, and we acknowledge the fact. The problem really amounts to a question of whether we should develop skills that are indeed interrelated or search for skills that are discrete. Actually, we have had little difficulty arriving at an answer to that question. We are convinced that the most straightforward way to help children become independent in tackling a diversity of comprehension tasks is to help them sharpen specific skills that are related to the process of comprehending. For purposes of efficient measurement, discrete subtests (subskills) may be desirable. But for purposes of coherent teaching, development of the overall process by sharpening related behaviors seems sensible.

IDENTIFYING SENTENCE MEANING SKILLS

KARLYN KAMM
University of Wisconsin–Madison

ABSTRACT

The paper describes the Sentence Meaning skills of an objective-based approach to reading comprehension. The assumption underlying this strand of skills is that getting meaning from a sentence, or a complete thought, is an integral step in the reading process.

The Sentence Meaning strand is developed through two sets of skills, one dealing with analysis and one with synthesis. There are six levels of development with either one or both sets of skills presented at each level. The organizing thread, or vehicles for moving through the various levels of skills is the **sentence structure.** At each level the child applies the skill (s) in the context of various structural specifications, which are generally more complex at each subsequent level. These specifications were identified for instructional emphasis on the basis of current research and the developers' experience with pilot testing of comprehension materials.

The analysis skills focus on the details of a sentence and require the learner to attend solely to a single detail within a sentence. These necessary basic skills were included since too frequently they are overlooked in developing learners' comprehension skills. The synthesis skills involve the learner in demonstrating his understanding of a sentence by requiring him to paraphrase given structures. The learner restates by both rearranging words and by substituting words.

IDENTIFYING PASSAGE MEANING SKILLS

GINGER VAN BLAIRCOM
University of Wisconsin–Madison

SANDRA WHITE
University of Wisconsin–Madison

ABSTRACT

As part of the Comprehension element of a skill-based approach to reading instruction, skills related to passage meaning have been developed. The skills in passage meaning are divided into two parts or strands. One strand focuses on determining the central thought of a passage while the other strand emphasizes understanding the relationships in a passage and forming conclusions based on these relationships. Central Thought skills deal with the ability to determine what the whole passage is about. Relationships/Conclusions skills deal with the ability to identify and understand information given and then to apply this information in forming conclusions. The purpose of this paper is (a) to give the rationale for including these two strands in passage meaning (b) to explain the present conceptualization of these two strands, and (c) to identify the skills in each strand.

IDENTIFYING WORD MEANING SKILLS

BEVERLY MORRISON
University of Wisconsin-Madison

ABSTRACT

In determining the skills for an objective based approach to the development of reading comprehension abilities, it became apparent that some emphasis should be placed on the skills necessary to determine word meaning. The paper purposes skill strands in context clues and word parts which would ultimately enhance the ability of an individual to comprehend what he reads. A rationale to justify the inclusion of these skills in a reading comprehension program is presented. The strands of skills, sequentially though not hierarchically arranged, are discussed. The skills are presented as points of emphasis to assist the individual reader in determining word meaning. The research and developmental process pursued in the identification and delineation of these skills in objective form is described. Consideration is also given to the instructional implications of the proposal.

VARIABLES IN THE DEVELOPMENT OF MATURE READING STRATEGIES

PHIL L. NACKE
University of Kentucky

JAMES A. CULBERTSON
Kentucky State University

The term mature strategies seems to be important because it is more comprehensive and integrative than the term skills. Flexible, efficient reading has received a great deal of lip service during the last several decades, but researchers still seem to be investigating isolated skills without following up with the study of the integration of these so-called skills in the complete mature reading process as observed in the model adult reader when performing realistic (reality-based) reading tasks in common reading situations.

In fact, it seems that this preoccupation with isolating reading skills, whose measurement and function are questionable, has obscured the fact that the ultimate goal of reading is the functional integration of such skills or subskills. The mature reader does not normally perform tasks in which the subskills are used in isolation. Rather, the effective and efficient adult reader utilized the subskills in association and compliance with one another.

A major reason for this state of affairs can be traced to the traditional approach to curriculum development whereby decisions are made by determining the skills to be taught in the preschool, primary, middle school, and then the junior and senior secondary school, in turn. This has resulted in curricula which include many tasks which are strictly school-based activities which students will not perform outside of the school environment. A meaningful alternative is to identify the range of tasks which competent adults need or want to perform. These tasks can then be organized hierarchically in order of complexity so that the instructional program can focus on the appropriate levels of tasks. Following this approach of using the mature reader as the model, there is much less likelihood of involving the student in learning subskills which are unproductive and inappropriate. The mature reader model should also guide the researcher in identifying the significant variables in the mature reading process which should be investigated.

There is that body of research, on the other hand, which is so global in its design that it lacks real significance. It seems remarkable that just last month the editor of a familiar journal published a study of a simple pre-test/post-test programatic sort using a standardized test.

Early (1976), in discussing research in reading and writing, drew the following significant conclusions:

> While psycholinguistic studies try to determine what's happening when students comprehend, other researchers have been investigating the conditions under which students comprehend certain subjects (always measuring the intake by the output). Such studies have not proved very fruitful, because they have not defined carefully what was being measured, because they pitted one complex teaching method against another without being able to sort out the variables, or because they evaluated specific material rather than generalizable principles. Future studies will make greater use of observational techniques, interviews, introspection, and laboratory simulations employing computers and recording devices.

Some of these issues will be considered in greater depth below.

Time Factor.

In studying mature reading strategies we need to give more careful attention to the variable of reading *time*. The time factor can be handled either as a time-amount condition or as a time-limit condition. It seems that the time factor has typically been handled as a time-amount (as in studies by Calfee & Jameson, 1971; Coke, 1974; Smith & Kulhavy, 1974) or has been ignored. The time-limit condition seems to have considerable potential for research, as well as for instruction, if it is treated as a criterion variable. In this situation (time-limit), the amount of information to be recalled can, of course, be manipulated as, "How much important information can you get from this amount of reading material in _____ minutes or seconds?" The specified amount of time must be carefully and realistically manipulated, taking into account the basic level of the subject's performance and the type of discourse as well as the level of response (comprehension) expected. Psychological factors also need to be considered, for, if the time limit is set too low, the subjects may react in a negative manner and not even attempt to perform the task.

Carver (1975a) has discussed some additional factors that should receive attention when designing reading rate research. He has called for further study of the causes of between-individual differences in reading rate and the effects of within-individual changes in reading rate.

It should also be pointed out that a meaningful coordination of the strategy variable and the time variable can be achieved by manipulating

275

the manner of presentation of the stimuli passages through the amount of discourse presented on one page or visual exposure (i.e. where amount of time is controlled) and through graphic arrangements and manipulation of the printed discourse. Nacke (1970) and others have studied mature strategies by manipulating the printed format of the discourse which is presented in a strict time limit condition.

McConkie and Rayner (1973), for example, have used very sophisticated computerized research equipment to investigate significant aspects of the reading process from a strategy perspective. Thomas and Augstein (1976) have used a much less complex instrumentation to investigate the strategy variable. In such studies, the processing time has been productively investigated.

Carver (1975b) has developed a technique for manipulating attention during reading which he calls programmed prose. This procedure, applying his reading-input technique, involves manipulation of the stimulus material whereby, at every *N*th word interval, an incorrect word is added as an alternative to the correct word, requiring the subject to make a binary choice of the alternatives from the sense of the context. This is a variation in manipulating the stimulus material in order to study the strategy variable. The time factor can also be studied when using this technique.

Directions

The kind of directions which are given to the readers explaining the task to be performed is another important consideration. These directions either specify or suggest the strategy to be utilized.

A number of studies (A.C. Smith, 1963; H.K. Smith, 1967; Hill, 1968; Calby, 1971; McConkie, Rayner & Wilson, 1973) have indicated that readers do not normally adjust their reading in response to directions to read for different purposes. This can probably be explained by the years of training to read everything carefully and the fear of failure, along with the unrealistic expectation that readers should be able essentially to reproduce all the information that is given without consideration for such things as the amount of background information, the terminology used in the passage, the progression of ideas or the style of presentation.

This suggests that research and instruction in mature reading strategies should take into account the practical distinction between general reading and studying as in studies by Anderson (1974) and Killer (1973). Studying normally requires different strategies because of the need to understand and recall more detail and to be able to store and retrieve this amount of information over a longer period of time.

This distinction emphasized the importance not only of the directions given explaining the task to be performed and any constraints (such as the time and the amount and type of information to be stored) but it also suggests the necessity of giving careful attention to the number of trials required for a given task. It is imperative that such variables be dealth with systematically in both research and classroom instruction.

Stimulus Materials

Another major problem area in research seems to be the nature and type of stimulus passages used. Variations in the semantic and syntactic features of the stimulus passages may very well account for some of the conflicting results of studies to date. The manner in which messages are embedded in discourse seems to be a confounding or problematic area in research. The formulas used to measure the readability of passages have been seriously questioned (Klare, 1974-75). It may be productive to conduct series of studies using the same stimulus passages rather than using different passages for each study. What may be needed, then, is a collection of carefully selected passages at different levels of difficulty covering a range of topics and representing selected styles of discourse.

Peters (1975-76) used the Frayer model of concept attainment to modify the content of stimulus materials in order to investigate the effect of the methods of concept presentation in the stimulus materials. Such types of manipulation of stimulus materials may prove to be a productive way to develop batteries of stimulus passages in which the manner of presenting the content is controlled.

The level of difficulty of the stimulus materials must be controlled and/or manipulated when studying mature reading strategies and when providing instruction on reading efficiency. Carver (1975a) has recommended more precise measurement and control of the differences between the level of difficulty of the stimulus materials and the reading ability levels of the subjects.

Background Information.

An important related factor is the control of the background information that subjects have about the topic. It seems inadequate to handle this through traditional reliability procedures. Rather, a more appropriate and necessary procedure is to control background information through the research design using a pre-test/no pre-test approach so that the effect of the pre-test itself can be controlled. This necessarily requires a larger *N* in such studies. An alternative for controlling background information is to develop orginal passages in which new information is generated. There are obvious limitations here, since a reader would likely not have the normal benefit of related prior knowledge and there is frequently a degree of artificiality in such passages.

Response Tasks.

With regard to the response tasks in research and instruction, it seems that answering multiple-choice questions has severe limitations as a method of checking comprehension and retention. This type of recognition task does not seem to be adequate as a means of determining the way in which adult or mature readers are able to process messages. Multiple-choice tasks are almost entirely relegated to school-based experiences. Outside of school situations adult readers typically need to use a recall type of response following (or even during) the act of reading. Therefore, it seems logical that both research and instruction should give a great deal more attention to response tasks which require the reader to produce the language which is needed to re-present the author's message rather than merely recognizing the language already produced for the reader.

Conclusion.

The rationale that guides these ideas is that research and instruction in reading to learn should be focused not only on what the adult reader needs to be able to do in the school situation, but, more important, on what the adult reader needs to be able to do outside of school settings when he/she is performing as an independent, mature learner in a realistic, day to day situation. By looking at adult reading from this reality-based perspective, we will be better able to set realistic and educationally significant priorities both for research and instruction.

At the risk of over simplifying, there seem to be three basic dimensions to keep in mind regarding adult reading and study strategies. The first is the process of selecting or identifying the essential message in discourse and using efficient strategies for locating the semantic and syntactic cues signaling that essential message.

The second aspect is the strategy of meaningfully translating this essential message into the reader's own experiential and linguistic frame of reference.

The third key aspect of adult reading really moves more specifically into study strategies. These are the rehearsal strategies that are necessary for confident and efficient storage and retrieval (or utilization) of the message that the reader selects and translates.

These aspects point to several dimensions of the mature reading process which need urgent attention: selecting and categorizing the common types of messages used for stimulus passages, analyzing the functional tasks required of readers and determining priorities for these tasks, identifying the problems readers have in processing messages within given time frames, and developing ways of providing direct instruction aimed at helping readers perform specific, practical reading tasks optimally.

References

Anderson, R.C. and others. Retention of text information as a function of the nature, timing, and number of quizzes. U.S., Educational Resources Information Center Document, 1974, No. ED089200.

Calby, D.H. The effects of specific directions on the reading compulsion of sixth-grade students. U.S., Educational Resources Information Center Document, 1971, No. ED 050905.

Calfee, R.C. and Jameson, P. Visual search and reading. U.S., Educational Resources Information Center Document, 1971, No. ED050924.

Carver, R.P. Designing reading rate research. In G.H. McNinch and W.D. Miller (Eds.), *Reading: Convention and Inquiry*. 24th NRC Yearbook. Clemson, S.C.: National Reading Conference 1975a, 241-245.

Carver, R.P. Manipulation of attention during reading using programmed prose. *Journal of Reading Behavior*, Fall 1975b, 7(3), 241-254.

Coke, E.U. Reading rate readability and variations in task-induced processing. U.S., Educational Resources Information Center Document, 1974, No. ED 092902.

Early, M. Important research in reading and writing. *Phi Delta Kappan*, January 1976, 298-301.

Hershberger, W.A. and Terry, D.F. Complexity of typographical cueing in programmed and conventional texts. Technical Report No. 7. U.S., Educational Resources Information Center Document, 1963, No. ED 094330.

Hill, W.R. Influence of direction upon reading flexibility of advanced college readers. In E.L. Thurston and L.E. Hafner, *New Concepts in College-Adult Reading*. 13th NRC Yearbook. Milwaukee: National Reading Conference, 1964, 119-125.

Killer, J.H. Effectiveness of various prose study techniques as a function of text readability, and individual differences. U.S., Educational Resources Information Center Document, 1973, No. ED 084492.

Klare, G.R. Assessing readability. *Reading Research Quarterly*, 1974-75, *10*, 62-102.

Mackworth, J.F. Some models of the reading process: Learners and skilled readers. *Reading Research Quarterly*, Summer 1972, *7*, 701-733.

McConkie, G.W., Rayner, R. and Wilson, S.J. Experimental manipulation of reading strategies. *Journal of Educational Psychology*, 1973, *65*, 1-8.

McConkie, G.W. and Rayner, K. An online computer technique for studying reading: Identifying the perceptual span. In P.L. Nacke (Ed.), *Diversity in Mature Reading: Theory and Research*, Volume 1, 22nd Yearbook of the National Reading Conference, 1973, 119-130.

Nacke, P.L. Skimming strategy in reading as a function of familiarity with content and redundancy reduction in printed discourse. Unpublished doctoral dissertation, University of British Columbia, 1970.

Peters, C.W. The effect of systematic restructuring of material upon the comprehension process. *Reading Research Quarterly, 11*(1), 1975-76, 87-111.

Smith, A.C. The influence of change in purpose upon ocular motor reading behavior of university freshmen. (Doctoral dissertation, University of Oregon) Ann Arbor, Michigan: University Microfilms, 1963.

Smith, H.K. The response of good and poor readers when asked to read for different purpose. *Reading Research Quarterly,* Fall 1967, *3,* 53-83.

Smith, P.E. and Kulhavy, R.W. The influence of adjunct rules and objectives on learning from text material. U.S., Educational Resources Information Center Document, 1974, No. ED 090508.

Thomas, L.F. and Augstein, E.S. Reading to learn. *New Horizons in Reading.* Newark, Delaware: International Reading Association, 1976, 147-157.

THE RELATIONSHIP OF THE UNIQUENESS OF PROSE PASSAGES TO THE EFFECT OF QUESTION PLACEMENT AND QUESTION RELEVANCE ON THE ACQUISITION AND RETENTION OF INFORMATION

MARK G. RICHMOND
University of Southern Mississippi

Objective of the Inquiry

Since 1966 much research has been conducted to investigate the placement of questions within textual material as a facilitative agent to retention. Rothkopf (1966, 1970) has termed studies of this genre as **mathemagenic**, or those which gave birth to learning. Frase (1967, 1968a, 1968b, 1968c has been the other major contributor to the data base developed around the mathemagenic concept. While others have done research in the area, the Frase and Rothkopf research is considered the most important for the following reasons:

1. Over 85 percent of the studies done have been patterned after, and are general replications of the research conducted by Frase and/or Rothkopf.
2. This secondary body of research makes use of the passages and/or questions developed by either Frase or Rothkopf for use in their studies.
3. The Frase and/or Rothkopf studies have defined the parameters of the studies in the areas of questions placement for enhancement of retention.

The parameters which Frase and/or Rothkopf have defined have generally dealt with the following areas:

1.0 Placement of questions:
  1.1 Questions before (QB): questions are placed immediately before the segment of the passage from which information is to be retained.
  1.2 Questions after (QA): questions are placed immediately after the segment of the passage from which information is to be retained.
  1.3 No questions (NQ): no questions are embedded in the text.
2.0 Relevance of questions:
  2.1 Relevant questions (RQ): questions on the retention measure (posttest) are deemed relevant if they were embedded in the text.

2.2 Incidental questions (IQ): questions on the retention measure (posttest) are deemed incidental if they appeared on the retention measure, but were not embedded in the text.

As previously mentioned, the same treatment passages and retention measures were used in repeated experiments purporting to investigate various mathemagenic factors. Not only did the primary investigators repeatedly utilized the same instruments, but generally the same instruments were used in ancillary investigations (Sanders, 1970; Cunningham & Keller, 1972).

Coleman (1964) points up the fallacy of over-relying on a single language sample for purposes of generalization. The fact that only two samples were employed in repeated investigations and that the results obtained in using the two instruments were never compared appears to have been a severe weakness in previous question placement research. Further, no attention appears to have been given to structuring the formation of questions and/or distractors in order to ensure to any degree the reliability of measurement. Bormuth (1970) has suggested a taxonomy for developing comprehension question stems, while Schlessinger and Weiser (1970) outlined a paradigm for distractor development. From this background, this study was conceived in line with the following purposes: (a) to investigate the differential effects of passage uniqueness on the outcome of mathemagenic studies by employing samples from a defined language population; and (b) to formalize the structuring of comprehension questions so that development of instruments would have a degree of consistency from one treatment passage to the other.

Subjects and Instruments

The subjects for the study were 596 sixth-grade students attending Meridian Middle School in Indianapolis, Indiana. Achievement test data (from Iowa Test of Basic Skills) indicated the composite grade level average of the sample was 6.2, when measured in the first month of the school year. Subjects were assigned to treatments in a totally randomized fashion. Treatment administration for all subjects was carried out simultaneously with a subject: proctor ratio of 18:1.

Three passages (A, B, and C) of approximately 600 words were randomly drawn from one source (Gray, 1942) and were equated for readability, with reading difficulty levels varying from a low of 5.14 to a high of 5.55. Using the three passages, treatment instruments were constructed by embedding questions either before (QB) or after (QA) portions, or by not embedding questions (NQ). For each passage, two questions were developed for every one question embedded in the QB and QA condition. On a random basis, one-half of the questions were assigned to the relevant questions condition with the paired question being assigned to the incidental questions condition by default. Thus, the posttest consisted of two distinct question types: (a) relevant questions, those which appeared in the text and on the posttest, and (b) incidental questions, those which appeared only on the posttest.*

Data analysis

Data were analyzed using two separate designs. Design 1 was a 3 x 3 x 2 x 2 mixed effect ANOVA with repeated measures on the D factor. Design 2 was a 3 x 2 x 2 x 2 mixed effect ANOVA with repeated measures on the D factor. Factors in both designs were:

A. Passage [P]
B. Question placement (QB, QA, NQ) [Q]
C. Immediacy (testing immediately after reading; testing delayed two weeks) [I]
D. Question relevance (posttest questions embedded in text, posttest questions not embedded in text) [R]

Design 2 was implemented because of a basic question of logic concerning the question placement factor. As no questions were included in the text for the NQ level of that factor, all posttest questions fit the operational definition of incidental questions. To determine if the logical difference was real, it was decided to repeat the data analysis but exclude the NQ data from the second analysis.

Results

Analysis of variance under design 1 revealed significant ($p < .05$) main effect differences among passages and between immediacy of testing, a significant ($p < .05$) first order interaction between passages and relevance, and two significant ($p < .05$) second order interactions: passages x placement x relevance and passages x immediacy x relevance.

Simple main effects analysis of variance performed on the P x R interaction revealed that within passage A relevant and incidental questions differed in retention producing strength, with RQs producing the higher retention score. Simple interaction effects analysis of variance for the P x Q x R interaction indicated that significant Q x R interactions occurred in passage A and in passage B but not in passage C. Interestingly, simple interaction effects analysis of variance for the P x I x R interaction indicated a significant first order interaction (I x R) only in passage C.

In the second analysis, employing only two levels of question placement (QB and QA), significant differences in the passages and immediacy of testing factors were again noted. Results of analysis 2 differed from analysis 1 in that only two interactions were noted in analysis 2: the P x R interaction and the P x Q x R interaction. Simple main effects analysis of variance on the first order (P x R) interaction indicated similar results to analysis 1 with the only significant

Table 1
Analysis of Variance on Number of
Correct Test Items – Three Groups
(NQ, QB, QA)

Source	d.f.	MS	F
Between Subject			
Passages (P)	2	168.78[a]	26.56*
Question Placement (Q)	2	31.73[f]	2.61
Immediacy of Testing (I)	1	2437.07[b]	1074.64*
P x Q	4	12.14[c]	<1
P x I	2	2.27[a]	<1
Q x I	1	0.25[d]	<1
P x Q x I	4	7.66[d]	1.21
Within Subjects			
Question Relevance (R)	1	22.72[c]	<1
P x R	2	35.10[a]	15.18*
Q x R	2	2.11[e]	<1
I x R	1	5.60[d]	<1
P x Q x R	4	6.47[d]	2.80*
P x I x R	2	9.54[d]	4.13
Q x I x R	2	0.03	<1
P x Q x I x R	4	4.94[d]	2.14

Error Terms:	Source	d.f.	MS
	a. Subjects (S)	432	6.36
	b. P x I	2	2.27
	c. P x R	2	35.10
	d. S x R	432	2.31
	e. P x Q x R	4	9.54
	f. P x Q	4	12.14

* $p < .05$

interaction occurring in passage A. Simple interaction effects analysis of variance of the P x Q x R interaction again indicated that significant first order Q x R interactions occurred in passage A and passage B but not in passage C.

Table 2
Simple Main Effects Analysis of Variance For the P x R Interaction (Analysis 1)

Source	SS	df	MS	F
P x R	70.20	2	35.10	15.18*
R at P_A	89.65	1	89.65	36.39*
R at P_B	3.00	1	3.00	1.22
R at P_C	0.27	1	0.27	< 1
MS pooled		239	2.46	

*$p < .05$

Table 3
Simple Interaction Effects Analysis of Variance for the P x Q x R Interaction (Analysis 1)

Source	SS	df	MS	F
P x Q x R	25.89	4	6.47	2.80*
QR at P_A	17.39	1	17.39	7.40*
QR at P_B	10.86	1	10.86	4.26*
QR at P_C	1.86	1	1.86	< 1
MS posted		424	2.35	

* $p < .05$

Table 4
Simple Interaction Effects Analysis of Variance For The P x I x R Interaction (Analysis 1)

Source	SS	df	MS	F
P x I x R	19.08	2	9.54	4.13*
IR at P_A	1.93	1	1.93	<1
IR at P_B	3.00	1	3.00	1.28
IR at P_C	19.77	1	19.77	8.43*
MS pooled		416	2.35	

* $p < .01$

Table 5
Analysis of Variance on Number of Correct Test Items—Two Groups (QB, QA)

Source	d.f.	MS	F
Between Subjects			
Passages (P)	2	113.51^a	16.69*
Question Placement (Q)	1	0.0^f	< 1
Immediacy of Testing (I)	1	1600.67^b	459.74
P x Q	2	7.09^c	1.04
P x I	2	3.48^a	< 1
Q x I	1	0.24^d	< 1
P x Q x I	2	3.82^d	< 1
Within Subjects			
Question Relevance (R)	1	25.63^c	1.05
P x R	2	24.33^a	10.69*
Q x R	1	0.11^c	< 1
I x R	1	3.23^d	< 1
P x Q x R	2	12.87^d	5.66*
P x I x R	2	3.92^d	1.72
Q x I x R	1	0.0^d	< 1
P x Q x I x R	2	4.64^d	2.04

Error Terms:	Source	d.f.	MS
	a. Subjects (S)	288	6.80
	b. P x I	2	3.48
	c. P x R	2	24.33
	d. S x R	288	2.28
	e. P x Q x R	2	12.87
	f. P x Q	2	7.09

* $p < .05$

Table 6
Simple Main Effects Analysis of Variance For the P x R Interaction (Analysis 2)

Source	SS	df	MS	F
P x R	48.66	2	24.33	10.69*
R at P_A	72.00	1	72.0	29.65*
R at P_B	0.85	1	0.85	<1
R at P_C	1.45	1	1.45	<1
MS pooled		185	2.43	

*$p < .01$

Table 7
Simple Interaction Effects Analysis of Variance For the P x Q x R Interaction (Analysis 2)

Source	SS	df	MS	F
P x Q x R	25.74	2	12.87	5.66*
QR at P_A	15.68	1	15.68	6.67*
QR at P_B	10.13	1	10.13	1.31*
QR at P_C	0.05	1	0.05	<1
MS pooled		255	2.35	

*$p < .05$

Discussion

In previously conducted research surrounding the mathemagenic hypothesis, the effect of variant passages upon the outcomes had not been investigated. Generally, invistigations have centered on question placement within text and on question relevance. Previous studies have generally reported superiority of QA treatments over QB or NA treatments. Further, previous research has indicated that in the instance of RQs vs. IQs, QAs appear to have a greater facilitative recall effect on IQs, while QBs exert strongest influence on RQs. Among various researchers, results have not been highly consistent. The lack of consistency has two possible explanations:

1. Lack of control in construction and use of materials. As Coleman (1964) has pointed out, most comprehension studies lack generalizability. This lack of generalizability stems from a lack of rigor in dealing with prose. Coleman strongly suggests that passages be selected under the same conditions (with the same cautions) as are subjects in a research study. Further, multiple passage studies should include the passage effect as an independent source of variance.

2. Variances of test scores is not under the control of the passage. Alternatively stated, comprehension questions used may lack, totally or in part, passage dependence. In virtually all question placement studies, there has been a noticeable lack of reliability data, and nowhere has pilot research been conducted to ensure passage dependence of items.

The two cited explanations probably account for much of the inconsistency of previous question placement studies. Each must be taken into account.

The findings of the present study seem to shed some light on the ambiguity of prevous research. While significant main effects of passages and immediacy of testing were found in both analyses **F** (2,423) = 26.56, **p** $<$ .05 and **F** (2,288) = 16.69, **p** $<$ .05 respectively for passages; **F** (1,2) = 1074.64, **p** $<$ .05 and **F** (1,2) = 459.74, **p** $<$ .05 respectively for immediacy), these main effects were not the focal point of the study. The main effect of passages can be explained by the fact that passages paired with questions were not of equal difficulty. However, for "mathemagenic theory" to be effective, the passage must maintain a consistency with the other factors.

A significant main effect of immediacy of testing is highly predictable. With an interval between testing of two weeks, much of what was read could have been forgotten. Conversely, if no significant main effect of immediacy had been found, such results would have been taken as an indication of passage independence of test items.

It is of interest to note that no significant question placement effect and no significant question relevance effect were demonstrated in the current study. In previous research either or both of these effects were present to some degree, and as a general rule, interactions were largely ignored. Such a practice quite probably led to overemphasis on main effects.

The interaction of the passages factor and other factors of this study are where the central research questions were answered. The passages x relevance interaction indicated that differences between relevant and incidental items were not consistent across passages. A parsimonious explanation for this finding is that passages contribute a unique variance to the relevance factor, thus not allowing for a mathemagenic explanation. Refer to table 2 and 6 for results of differences between passages of relevance of questions. The tables indicate that in analysis 1, relevant and incidental questions gave significantly different subtest scores for passage A, but that there were no significant differences in relevant and incidental items in passage B or C. In analysis 2, the same pattern is noted.

The interaction between question placement and question relevance across passages is displayed in tables 3 and 7. Both analyses indicated that while a significant Q x R interaction existed in passage A and in passage B, no such interaction between question placement and question relevance was found in passage C. Again, such results cannot be explained in mathemagenic terms.

A third interaction was noted in analysis 1, but was not found in analysis 2. The interaction of immediacy of testing with question relevance across passages evidenced the same lack of consistency across the three treatment passages. A significant I x R interaction was asked for passage C but was not evident in either passage A or passage B. Again, it appears that differences between passages are equally as important as treatment conditions in studies of verbal comprehension.

Coleman (1964) has warned of the dangers of extrapolating to a language population without adequate selection and analytic safeguards. It seems warranted on the basis of this study to urge strongly that future research in the area of comprehension include a passages factor for which the passages have been carefully selected. In the current study, although readability was controlled (to within a spread of 5 months), questions were carefully constructed and pilot tested, and tests evidenced substantial reliability, the passages themselves did not allow a consistent pattern of effects to be noted through the study.

The analysis question (problem statement 3) appears to be answered in two different respects. The primary difference between the two analyses is that in analysis 1 there was a significant P x I x R interaction while no such interaction was found in analysis 2.

The fact that an additional significant interaction occurred when the NQ level of question placement was included in the analysis appears to indicate that treating non-embedded questions as relevant forced a greater amount of variance to occur across the question relevance factor in the current study. Because of the apparent artificiality of this type of variance, it would seem highly preferable to use a design which did not employ an NQ treatment when investigating both question placement and question relevance. In the present study the only rationale for inclusion of analysis 1 was to facilitate comparison of the results of this study with previous research in which an analysis comparable to analysis 1 was used. Analysis 2, however, provides the most logical interpretation of the data.

A surprising outcome of the conducted research was the absence of a significant difference on the question placement factor. In virtually all of the reviewed research, a significant difference was found to exist among the three levels of question placement; however, this previous research did not allow predictions as to directions of the differences. Older studies (before 1960) tended to report that questions placed after relevant portions of text aided in later recall of information or a posttest to a greater degree than questions placed before relevant portions of text. More recent research (since 1960) has tended to produce results in the opposite direction, that is, finding that questions placed before relevant text increase posttest scores to a significantly greater degree than do questions placed after relevant text. In addition, in many studies treatments employing no questions embedded in passages were found not to be different from one or the other of the placements, with the NQ treatment scores occasionally exceeding either QB or QA treatments. The primary point of interest, however, is that, generally, significant differences were reported in posttest scores attributable to question placement.

The lack of such a significant effect in the present study cannot be overlooked. It would appear that there are three possible explanations for the lack of a significant placement effect:

1. Psycholinguistic research has indicated that information received in input as linguistic stimulus is retained in long term memory in semantic form, but the lexical form of the input is not retained. The embedded questions as presented in the current research studies were designed to elicit responses which were not only passage dependent but also dependent on the syntactic environments from which they were drawn. Since subject may have been required to draw upon lexical rather than semantic memory, the answers to the test questions could have necessarily been drawn to a large degree from short term memory. With this factor present, it would appear that variance among treatment groups on posttest scores could have been homogenized due to functions of memory, rather than functions of question placement:

2. Similarly, memory research has shown that questions placed in time proximity too close to stimulus material do not aid in storing content in long term memory. As questions in the current study were in very close proximity with the content from which they were drawn, it

is possible that QBs and QAs were no more effective than NQs in aiding the storage of information contained in a given passage.

3. A third possible explanation for the unexpected lack of differences among treatment groups deals with the specificity of the questions. Questions were constructed in such a manner that their content was totally tied in with the passage. This being the case, it would seem that whether or not questions were embedded or the position in which they were embedded had little effect on whether content was retained. The retention of information within a passage under such circumstances would be highly dichotomous; either information was retained or it was not, without regard to question presence or position.

The three explanations provided above are important aspects of the present study and should be of some interest to psychological researchers in the areas of linguistics and memory.

Selected References

Bormuth, J.R. *On the theory of achievement test items.* Chicago: University of Chicago Press, 1970.

Coleman, E.G. Generalizing to a language population. *Psychological Reports,* 1964, **14**, 219-226.

Cunningham, D.J., & Keller, D.F. *Recall instructions and learning from text with adjunct questions.* Research Report. Bloomington: Indiana University, Institute for Child Study, 1972.

Frase, L.T. Learning from prose material: Length of passage, knowledge of results, and position of question. *Journal of Educational Psychology,* 1964, **58**, 266-272.

Frase, L.T. The effect of selected variables upon the retention of prose material: Study II. *Journal of Educational Psychology,* 1968(a), **59**, 244-249.

Frase, L.T. *Effective prose reading: Shaping and discriminative effects of questions.* Paper presented at the meeting of the American Educational Research Association, Chicago, 1968(b).

Frase, L.T. Some data concerning the mathemagenic hypothesis. *American Educational Research Journal,* 1968(c), **5**, 181-189.

Gray, E.J. *Adam of the road.* New York: Viking Press, 1942.

Rothkopf, E.Z. Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal,* 1966, **3**, 241-249.

Rothkopf, E.Z. The concept of mathemagenic activities. *Review of Educational Research,* 1970, **40** (3), 325-336.

Sanders, J.R. *Short and long-term retention effects of adjunct questions in aural discourse: An extention os research on mathemagenic behavior.* Unpublished doctoral dissertation, University of Colorado, 1970.

Schlessinger, I.M., & Weiser, Z. A facet design for tests of reading comprehension. *Reading Research Quarterly,* 1970, **5** (4), 556-580.

A MODEL OF THE COMPREHENSION OF POETRY

W. JOHN HARKER
University of Victoria

Recent models of the reading process have been largely concerned with applying psycholinguistic knowledge to the reading of prose (Goodman, 1970; Ruddell, 1970; Smith, 1971). In these models, reading prose is represented as information processing. The mature reader is characterized as an active user of language in his search for meaning. He is seen to mobilize his sensory, perceptual, and cognitive processes in a purposive endeavor to decode the meaning communication by the writer. He does this through a process of reducing his uncertainty (Smith, 1971) about the message, through making predictions and testing hypotheses based on his knowledge of language and the way language represents reality. Goodman (1970) has described the process as follows:

> Reading is a selective process. It involves partial use of available minimal language cues selected from perceptual input on the basis of the reader's expectation. As this partial information is processed, tentative decisions are made to be confirmed, rejected, or refined as reading progresses. (p. 260)

When the reader turns his attention to poetry, however, he encounters reading material quite different from prose. Poetry is more complex, more abstract and therefore more difficult to comprehend than prose. In discussing meaning in poetry, T. S. Eliot (1942) has written:

> If . . . only a part of the meaning can be conveyed by paraphrase, that is because the poet is occupied with frontiers of consciousness beyond which words fail, though meanings still exist. (p. 15)

Similarly, Wimsatt (1954) has described a poem as:

> . . . a complex kind of verbal construction in which the dimension of coherence is by various techniques of implication greatly enhanced and thus generates an extra dimension of correspondence to reality, the symbolic or analogical (p. 241)

The ideational complexity of poetry has also been pointed to by Wellek and Warren (1949) who have described a poem as "a highly complex organization of a stratified character with multiple meanings and relationships" (p. 17). And finally, Tate (1936) has observed that "in poetry all things are possible . . . because in poetry the disparate elements are not combined in logic . . . " (p. 335).

Poetic Language

It is clear from these descriptions of poetry that the reading of poetry, like the reading of prose, involves the cognitive processing of language, but laguage of a unique kind. In poetry, to a greater degree than in prose, language and meaning become fused in a unified artistic whole. In this way, a poem takes on an existence of its own which transcends the language in which it is written. Stankiewicz (1961) has described this phenomenon as "interiorization" in which the language of a poem creates "a codified structure" (p. 15). Similarly, Levin (1962) has observed that, through its language, a poem "generates its own code, of which the poem is the only message" (p. 41).

The Grammar of Poetry

The nature and function of poetic language become clearer when the grammar of poetry is examined. Slobin (1971) has noted that "much of the understanding of certain kinds of poetry is based on . . . the ability to find interpretation for grammatically unusual constructions" (p. 5). For example, when the reader encounters the line, "a grief ago," from Dylan Thomas, it is obvious that the grammar of Standard English has somehow been altered.

The relationship between poetic grammar and the grammar of Standard English has received the attention of a number of linguists. Chomsky (1961) has considered poetic utterances in terms of their relative "degrees of grammaticalness." Katz (1964) has developed the notion of "semi-sentences," while Ziff (1964) has explained poetic discourse in terms of its "deviance" from Standard English. Stankiewicz (1960) has observed that the study of poetic language consists of "a study of a certain type of rearrangement and modification of the elements of everyday spoken language" (p. 70). Harris (1952) has described the language of poetry as "the dialect within which the text was spoken or written" (p. 4), while Jakobson (1960) has referred to "the grammar of poetry" (p. 375). Through all of these descriptions, the relationship between poetic language and Standard English is represented as one of degrees of deviance from the norm rather than clear-cut difference. Although poetic utterances display grammatical rules which do not conform to Standard English, the rules of poetic grammar are related to the rules of Standard English.

Rather than to obscure meaning, however, the effect of grammatical deviance in poetry is to enrich and focus meaning. The fact that the grammatical rules of Standard English may be violated provides the poet with almost unlimited linguistic opportunities to infuse meaning in his poem. For example, returning to Thomas' line, "a grief ago," Levin (1964) has shown that by lifting grammatical restrictions, "grief" becomes associated with the subclass of temporal nouns while retaining its association with a state of mind. Thus the two meanings are fused. This deliberate and controlled ambiguity serves to enrich the meaning of the poem rather than to obscure it.

This example serves to illustrate that grammatical deviance is an integral function of poetic discourse, not an accident or contrivance. The poet as an encoder of language finds the language to be inadequate to express his meaning; he therefore alters the language and in doing so violates its rules to produce deliberate ambiguity. In support of this notion, T. S. Eliot (1942) has observed that "the ambiguities [in a poem] may be due to the fact that the poem means more, not less, than oridinary speech can communicate" (p. 16). Literary critics (Empson, 1947; Wimsatt, 1954) have demonstrated that linguistic ambiguity provides the basis for much of the aesthetic intensity and ideational complexity of poetry.

The Effect of Poetic Language

The effect of this grammatical deviance in poetry is to frustrate expectancy. Psycholinguistic research indicates that the mature reader of prose makes use of the redundancy in language to gain meaning. The reader's experience in learning and using Standard English have developed within him an expectancy for meaning to be revealed through predictable grammatical structures (Miller, 1962). As Smith (1971) notes, "a reader rarely appears to be delayed in identifying a word because he is considering that it might be something else" (p. 138). The reader does not expect ambiguity; his entire experience as a language user has taught him that language 'makes sense,' and its sense can largely be anticipated and predicted. This contention is supported by Kolers' (1970) research which shows that the reader is more sensitive to the grammatical relationships among words than he is to the words themselves.

But for the reader of poetry, much that has come to be expected from language is lost. T. S. Eliot (1933) has described his plight:

> . . . there is the difficulty caused by the author's having left out something which the reader is used to finding; so that the reader, bewildered, gropes about what is absent, and puzzles his head for a kind of 'meaning' which is not there . . . . (p. 144)

For the meaning of poetry to become clear, the reader must somehow hold in abeyance much of what he has learned to expect from Standard English while at the same time learn what to expect from the language of poetry. He must in effect recategorize what he knows about language in order to accommodate what he seeks to know from the poem. Bruner (1956) has described this process, observing that "the effectiveness of poetry often rests on its ability to cut across our conventional linguistic categories in a way evocative of more affective categories" (p. 5). Seen in this light, reading a poem becomes analogous to learning a new language, to developing new linguistic intuitions about its structure. Put another way, the reader comprehends a poem to the extent that he has learned the new grammatical rule system of poetry.

Parallelism

The rules of poetic language, though deviant from Standard English, are nevertheless rules and they must be learned for poetry to be comprehended. Although, as Chomsky (1961) has noted, the comprehension of poetry involves discoveirng analogies between the structure of poetic discourse and Standard English, learning the rules of poetic discourse involves considerably more than analogizing. This is because the grammatical deviance apparent in poetry is not random; it is rule-governed and has a structure of its own which gives meaning to the poem and therefore provides the reader with a key to the poem's comprehension. As Mukarovsky (1964) has put it:

> The violation of the norm of the standard [language], its systematic violation, is what makes possible the poetic utilization of language; without this possibility there would be no poetry. (p. 42)

The systematic violation of the grammar of Standard English, the sustained patterning of grammatical deviance within poetry, is a generally recognized phenomenon. Gerard Manley Hopkins, certainly one of the most "ungrammatical" English poets, wrote that "the artifical part of poetry, perhaps we shall be right to say all artifice, reduces itself to the principle of parallelism" (Jakobson, 1966, p. 399). This concept of parallelism Jakobson has described as a process of "recurrent returns" (p. 399).

It is largely through the mechanism of parallelism that grammatical deviance in poetry established its own structure which then becomes the grammar of poetry. Halliday (1964) has described this process as one through which the poet "finds and exploits" irregularities in language, and in doing so, "superimposes a further regularity" (p. 305). While not all parallelism in poetry is structured according to grammatical deviance (Leech, 1969, p. 56 ff), grammatical deviations in poetry only achieve poetic effect when, as Bierwisch (1970) has noted, "the deviation has a specific regularity as its basis, when they [sic] stop being merely violations of the grammatical rules" (p. 110). Levin (1962), for example, has demonstrated how parallelism, or "coupling" to use his term, works within a Shakespearean sonnet through a process of pairing semantically equivalent elements set in equivalent syntagmatic positions.

Foregrounding

Despite the pervasiveness of parallelism in poetry, prolonged structural regularity leads to banality. Successful poetry overcomes banality by exhibiting what Mukarovsky (1964) has termed "foregrounding." Through foregrounding, the linguistic expectancy developed in the reader by the consistently structured grammatical deviance of the poem is deliberately violated to heighten meaning. Without this violation, the consistent structure of grammatical deviance would itself become normalized and the meaning obscured through banality. For example, Leech (1969, p. 58) has shown how in *The Waste Land*, T. S. Eliot uses fragments of colloquial, converstational language against a background of poetic language to focus meaning.

The effect of foregrounding on the reader is surprise and heightened attention to meaning. As Brown (1960) has observed, "the resulting combination of a strong expectancy and an improbable event communicates two meanings rather than one" (p. 384). In this way renewed ambiguity and the resulting enrichment of meaning are provided by foregrounded linguistic features being imposed upon a regularized pattern of deviance. It is through this interaction of foregrounded and unforegrounded linguistic features that the overall poetic structure is generated. Again, to quote Mukarovsky (1964):

> The mutual relationships of the components of the work of poetry, both foregrounded and unforegrounded, constitute its **structure**, a dynamic structure involving both convergence and divergence and one that constitutes an undissociable artistic whole, since each of its components has its value precisely in terms of its relation to the totality" (p. 22).

Short-Term Memory and Poetry

The nature of poetic language places a heavy burden on short-term memory functioning during the reading process. The function of the short-term memory is to hold incoming information while it is being processed (Miller, 1956; Slobin, 1971; Smith, 1971). However, Miller (1956) has shown that the short-term memory is limited to holding between 5 and 9 units of information at one time. But Miller also demonstrated that in language processing the reader is able to extend the capacity of the short-term memory by "chunking" verbal information into larger units. He does this on the basis of his intuitive expectancies for semantic and syntactic redundancy in language. In this way, the capacity of the short-term memory, though remaining limited, is expanded.

During the reading of poetry however, everything is working against the reader. The limited capacity of the short-term memory imposes an even more stringent bottleeck on information processing than it does during the reading of prose. The predictions the reader can make and the expectancies he can hold regarding the structure of language while reading prose are largely invalid when reading poetry. As has been demonstrated, the grammar of poetic discourse not only deviates from the grammar of prose, but poetry sets up its own grammar upon which it then proceeds to play variations. All this imposes a tremendous burden on the short-term memory. The reader unused to the language of poetry, finds the chunking of poetic language into larger units of meaning almost impossible. He is reduced to the position of the child who must carefully scrutinize every feature of the message in order to search for its meaning.

The Sound of Poetry

While restricting the capacity of the short-term memory, poetry provides aids to the reader which, while not entirely compensating for the difficulties imposed by its grammatical deviance, do assist the reader in the task of comprehension. The first of these aids to be considered is the sound patterning of poetry as developed through rhyme, meter, and figures of speech.

Sound and Meaning

The relationship between sound and meaning in poetry is widely recognized. Jakobson (1960) has written that poetry "a province where the internal nexus between sound and meaning changes from latent to patent and manifests itself most palpably and intensely" (p. 373). Similarly, Stankiewicz (1961) has observed that sound in poetry becomes "a carrier of meaning" through creating "a nexus of relations which are simultaneously ambiguous and necessary" (p. 16). The nature of these relationships has been illustrated by Hymes (1960) who, in a study of 20 sonnets by Keats and Wordsworth, found the dominant sounds within the poems to be concentrated within those words which express the themes of the poems. In another investigation, Masson (1963) traced the sound-meaning relationships within the line, "And frightful a nightfall folded a rueful a day," from Gerard Manley Hopkins' *The Wreck of the 'Deutschland'*; he found the sound patterning of the line to enforce the "crescendo of horror" (p. 71) suggested by Hopkins' words.

Sound as a Mnemonic Device

Sound patterning aids in the comprehension of poetry by serving as a mnemonic device (Charlesworth, 1975); the sound patterning of poetry expands the capacity of the short-term memory. Smith (1971) notes that information placed in the short-term memory is quickly lost if it is not constantly renewed by some form of internal rehearsal. Sound patterning in poetry constantly rehearses significant sound elements and concomitantly rehearses the meanings with which these elements are associated. This rehearsal allows the short-term memory time to hold information so that it may be processed before new information is introduced. The processing time of the short-term memory is also extended by pauses produced by meter and rhyme in poetry.

Closely associated with sound patterning is the function of parallelism. In his study of parallelism, Levin (1962) concluded that the effect of parallelism was intensified if rhyming words were semantically equivalent. Through a correspondence of meaning and sound, especially when sound becomes increasingly predictable through its parallel patterning in a poem, meaning also becomes increasingly predictable as redundancy develops. It follows that the linguistic units which the short-term memory is capable of chunking become larger as a function of their increased predictability. And as the units of meaning which can be held in the short-term memory become larger through more accurate prediction, comprehension becomes easier.

But, like everything else for the reader of poetry, the use of sound patterning as an aid to comprehension is not easy. Research (Brown, 1955; Garrett, Bever and Fodor, 1966) indicates rather conclusively that English language users attend to grammatical structures rather than to sound patterns during comprehension. In reading prose, the speed of the skilled mature reader precludes the possibility of "sounding" or subvocalizing the message, and strongly suggest that he goes directly to meaning (Smith, 1971, pp. 185 ff.). In this way speed aids comprehension. Thus it would seem that the reader of poetry must unlearn behavior which aids him in the comprehension of prose–speed–and slow down so as to attend to the sound of poetry. If this shift in attention does not take place, the aid to comprehension provided by the sound patterning of poetry will be lost to the reader as he is left to search for meaning at the grammatical level alone–and it has already been demonstrated how difficult this is.

The Shape of Poetry

Another aid to the comprehension of poetry is the configuration of the poem on the page, what Leech (1966) calls its "graphological deviation" (p. 47). Kell (1963) has shown how the line layout of poetry provides pauses which "vary both the timing and the shape . . . of the reader's thought" (p. 342). These pauses also allow for short-term memory rehearsal in the manner described in the previous section.

But the shape of a poem on the page may aid comprehension in another way as well. Kolers (1969) has described a study of the eye movements of a highly skilled speed reader which reveal fixations to be highly irregular and extremely selective. Kolers characterized this reader as follows:

> He has attained an information-transfer from the printed page to his own mind. It seems obvious that the transfer has been effected by the reader's constructing in his own mind a representation of what he has been reading about. This representation is thematic or semantic or information–no single word exists to describe it–and not literal. (p. 15)

This description can be applied to the process and result of reading a poem. The poem, through the physical arrangement of its linguistic features on the page, guides the eye to fixate on the loci of meaning within the poem–on those foregrounded linguistic features which charge the poem with meaning. In this manner, the poem's linguistic features and the meaning they convey become fused into what De Groot (1964) describes as the gestalt of the poem, "an organized whole or configuration of aesthetic experience" (p. 295). This process is particularly apparent in free verse such as the following example by E. E. Cummings:

seeker of truth

follow no path
all paths lead where

truth is here

No. 3 of **73 Poems**

Unfortunately, in order to take advantage of the physical configuration of poetry, the reader must ignore much of the behavior he has learned as a reader of prose–the rhythmatic left-to-right progression of the eyes across regularly-placed lines of print.

Summary and Conclusion

The model of comprehension of poetry which this paper presents is one in which the mature reader is constantly confronted with the unexpected. Much of what the reader has learned to predict about the structure of language is violated by poetry. Instead, poetry sets up its own grammatical rules and then proceeds to play variations on them in order to intensify the meaning communicated. In order to comprehend this meaning, the reader must abandon many of his expectations about language which can only interfere with his comprehension of poetry, and he must place himself in a position analogous to that of a beginning reader as he sets about learning the language of poetry. But even here, the reader must impose a further burden upon himself as he consciously decreases his reading speed so as to attend to the sound patterning of poetry as an aid to comprehending meaning. And even the, poetry does not arrange itself on the page in an expected and predictable manner. The reader must consciously look at and through and between lines in order to perceive the physical ordering of complex meaning in a poem. Through all this, the reader is required to alter and to bring to the conscious level the process of reading which for him has become largely unconscious while reading prose. It is this process of making conscious the reading act which distinguishes most clearly the reading of poetry from the reading of prose.

Any attempt to explain how art is understood, be it poetry or some other form, should of necessity end with an apologia. This attempt will be no exception. The model presented here is bound to be as over-simplified as it is tentative for the reason that there is wide variety among the reading processes of different readers and among the characteristics of individual poems. If there is any justification for seeking to explore the comprehension of poetry, it is because poetry represents the most exact and refined use of language by man, and through gaining understanding about how poetry is comprehended, we may learn something about both man and his language and the intricate relationship between the two.

References

Bierwisch, M. Poetics and linguistics. P. H. Salus (Trans.). In D. C. Freeman (Ed.), *Lingistics and literary style*. New York: Holt, Rinehart and Winston, 1970.

Brown, R. Opening statement: From the viewpoint of psychology. In T. A. Sebeok (Ed.), *Style in language*. Cambridge, Mass.: MIT Press, 1960.

Brown, R. W. Review. *Language*, 1955, 31, 84–91.

Bruner, J. S., Goodnow, J. & Austin, G. A. *A study of thinking*. New York: John Wiley, 1956.

Charlesworth, R. A. Don't forget poetry: Some observations on reading poetry. Paper presented at the Eighth Annual Reading Conference, York University, Toronto, February, 1975.

Chomsky, N. Some methodological remarks on generative grammar. *Word*, 1961, 17, 219–239.

De Groot, W. The description of a poem. In H. G. Lunt (Ed.), *Proceedings of the ninth international congress of linguists*. The Hague: Mouton, 1964.

Eliot, T. S. *The music of poetry*. Glasgow: Jackson, 1942.

Eliot, T. S. *The use of poetry and the use of criticism*. Cambridge, Mass.: Harvard University Press, 1933.

Empson, W. *Seven types of ambiguity*. (2nd ed.) London: Chatto and Windus, 1947.

Garrett, M., Bever, T. A., & Fodor, J. The active use of grammar in speech perception. *Perception and Psychophysics*, 1966, **1**, 30–32.

Goodman, K. S. Reading: A psycholinguistic guessing game. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading*. Newark, Del.: International Reading Association, 1970.

Halliday, M. A. K. The linguistic study of literary texts. In H. G. Lunt (Ed.), *Proceedings of the ninth international congress on linguistics*. The Hague: Mouton, 1964.

Harris, Z. S. Discourse analysis. *Language*, 1952, **28**, 1–30.

Hymes, E. H. Phonological aspects of style: Some English sonnets. In T. A. Sebeok (Ed.), *Style in language*. Cambridge, Mass.: MIT Press, 1960.

Jakobson, R. Closing statement: Linguistics and poetics. In T. A. Sebeok, (Ed.), *Style in language*. Cambridge, Mass.: MIT Press, 1960.

Jakobson, R. Grammatical Parallelism and its Russian facet. *Language*, 1966, **42**, 399–429.

Katz, J. J., & Fodor, J. A. The structure of a semantic theory. *Language*, 1963, **39**, 170–210.

Kell, R. A note on versification. *British Journal of Aesthetics*, 1963, **3**, 341–345.

Kolers, P. A. Reading is only incidentally visual. In K.S. Goodman & J.T. Fleming (Eds.), *Psycholinguistics and the teaching of reading.* Newark, Del.: International Reading Association, 1969.

Kolers, P.A. Three stages of reading. In H. Levin & J.P. Williams (Eds.), *Basic studies on reading.* New York: Basic Books, 1970.

Leech, G.N. *A linguistic guide to English poetry.* London: Longmans, 1969.

Leech, G.N. Linguistics and the figures of rhetoric. In R. Fowler (Ed.), *Essays on style and language.* London: Routledge and Kegan Paul, 1966.

Levin, S.R. *Linguistic structures in poetry.* The Hague: Mouton, 1962.

Levin, S.R. Poetry and grammaticalness. In H.G. Lunt (Ed.), *Proceedings of the ninth international congress on linguistics.* The Hague: Mouton, 1964.

Masson, D.I. Sound and sense in a line of poetry. *British Journal of Aesthetics.* 1963, **3**, 70-72.

Miller, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, 63, 81-92.

Miller, G.A. Some psychological studies of grammar. *American Psychologist*, 1962, 17, 748-762.

Mukarovsky, J. Standard language and poetic language. In P.L. Garvin (Ed. & Trans.), *A Prague school reader on aesthetics, literary structure, and style.* Washington: Georgetown University Press, 1964.

Ruddell, R.B. Psycholinguistic implications for a systems of communication model. In K.S. Goodman & J.T. Fleming (Eds.), *Psycholinguistics and the teaching of reading.* Newark, Del.: International Reading Association, 1969.

Slobin, D.I. *Psycholinguistics.* Glenview, Ill.: Scott Foresman, 1971.

Smith, F. *Understanding reading.* New York: Holt, Rinehart and Winston, 1971.

Stankiewicz, E. Linguistics and the study of poetic language. In T.A. Sebeok (Ed.), *Style in language.* Cambridge, Mass.: MIT Press, 1960.

Stankiewicz, E. Poetic and non-poetic language in their interrelation. In D. Davie et al (Eds.), *Poetics.* The Hague: Mouton, 1961.

Tate, A. *The man of letters in the modern world.* London: Meridian Books, 1957.

Thorn, J.P. Stylistics and generative grammars. *Journal of Linguistics*, 1965, 1, 49-59.

Wellek, R. & Warren, A. *Theory of literature.* London: Jonathan Cape, 1949.

Wimsatt, W.K. *The verbal icon.* Lexington: University of Kentucky Press, 1954.

Ziff, P. On understanding 'understanding utterances'. In J.A. Fodor and J.J. Katz (Eds.), *The structure of language.* Englewood Cliffs: Prentice-Hall, 1964.

A COMPARISON OF SYNTACTIC WRITING MATURITY WITH READING COMPREHENSION

NORMA JOHNSON
East Texas State University

ABSTRACT

The purpose of the study was to investigate the relationship between reading comprehension and 11 measures of syntactic writing maturity. These 11 measures were Hunt's: words per T-unit, subordinate clauses per T-unit, mean of the number of words per main clause, and mean of the number of words per subordinate clause; Golub's Syntactic Density Score; and Christensen's instances of free modification per T-unit, words in free modification per T-unit and per instances, instances of intra-T-unit coordination per T-unit, and words in intra-T-unit coordination per T-unit and per instance.

The experimenter requested 144 third-, fourth-, and fifth-grade children to write a story about a stimulus picture during a 30-minute time period. The next day the SRA Assessment Survey Achievement Series, Reading was administered by the teachers of each class.

Some significant low correlations occurred for the total group. Moderate coorelates (more than .40) occurred between the three reading scores and production of subordinate clauses for the 38 black students. Correlations were higher for the males than the females in words/T-unit.

A relationship between comprehension and writing muaturity was not clearly established. Some of Hunt's measures had low and moderate correlations with the reading scores; so as students wrote more words per T-unit, their reading level showed some increase. As black students added more subordinate clauses to their writing, their reading level increased. For whites, the reading growth indicator was words in free modification, a Christensen measure.

THE EFFECT OF STUDENT CONSTRUCTED QUESTIONS ON THE COMPREHENSION OF A PASSAGE OF EXPOSITORY PROSE

RONALD SCHMELZER
Anoka-Ramsey Community College

ABSTRACT

How to increase student learning of prose material has been a subject of discussion and research for over 50 years. Much of the research has concluded that students who question or have questions supplied to them while reading tend to comprehend more than those who have not questioned or haven't had questions supplied to them. The problem that was studied in this research was the effect of student constructed questions on comprehension. Four groups of Community college students (N=159) constructed questions in three different locations in a passage: (a) pre-questions, (b) interspersed questions, (c) post-questions, and one group (control) read the passage twice. Results suggested that post-questioning was significantly better than the other three methods for overall comprehension. However, when question type was analyzed pre-questioning and post-questioning were not significantly different, although both were better than interspersed questioning and reading twice.

MEASURING READABILITY AND READING ABILITY IN ORDER TO PREDICT COMPREHENSION[1]

RONALD P. CARVER
University of Missouri-Kansas City

The traditional approach to measuring readability and reading ability in order to predict comprehension needs to be improved. The traditional approach is to apply a readability formula to get a grade difficulty of the material and then to compare this to the grade level of ability of the individual, or group, as measured by a standardized reading test (e.g., see Cline, 1972). If the difficulty of the material (e.g., Grade Level 12) is higher than the ability of the individual (e.g., Grade Level 10), then the material is generally considered to be "too difficult," or that comprehension will be low or zero. If the reverse situation occurs, (i.e., the ability of the individual is higher than the difficulty of the material), then the material is generally not considered to be too difficult or that comprehension will be high. This article will deal with some of the problems associated with this approach when it is used to predict comprehension.

At the outset, it is evident that "too difficult" is a rather vague idea and that low or high comprehension is also rather imprecise. Conceptually, it seems more reasonable to consider percent comprehension as varying with the **difference** between the grade ability of the reader and the grade difficulty of the material. Then the research question becomes, what percent comprehension can we expect for each amount of grade difference? For example, when the ability of the individual is exactly equal to the difficulty of the material in grade level units can we expect 50% comprehension, 75% comprehension, or maybe even 100% comprehension? Or, if the difficulty of the material is one grade level higher than the ability of the individual, can we expect 50% comprehension, 25% comprehension, or maybe even 0% comprehension?

One reason why this question has not received much research attention is probably because percent of comprehension is difficult to measure accurately. Another reason for avoiding this type of research is that the percent of comprehension is not constant for a particular person and a particular passage, (e.g., the individual can increase the

---

[1] The preparation of this paper was supported in part by the National Institute of Education, Grant No. NEG-00-3-0043.

percent of comprehension by reading the passage more than once). In fact, this point is not always appreciated when making decisions about the mismatch between material difficulty and individual ability. If the material is three grade levels below the ability of the individual, then the individual might be expected to comprehend 75% or more upon one reading of the material. If the material is three grade levels above the ability of the individual, this may indicate that comprehension would be low upon the first reading, (e.g., 25%) but that by rereading the material several times, (e.g., concentrated study) then the individual could eventually achieve 75% or more comprehension. Therefore, instead of assuming that material is "too difficult" if it has a higher grade level of difficulty than the ability level of the reader, it could be that the material is not too difficult as long as the individual could be motivated to spend the additional time required to achieve an acceptable degree of comprehension, such as 75%.

Any attempt to predict comprehension from the match of the ability of the reader and the difficulty of the material should somehow control for time spent reading or studying. One way to deal with this problem is to try to predict comprehension in the following situation: the individual reads the material once at his normal reading rate.

Another problem with conducting research in this area is that the units used to measure difficulty are not necessarily the same as the units used to measure ability, even though both purport to measure along a scale of grades. The fact of the matter is that a grade ability score of 6 on a standardized reading test has not been defined so as to reflect anything about whether the student can read material at grade level 6 in difficulty; the score of 6 ordinarily means that this student answered the same number of questions correctly as the average of all the students in the normative group who were in grade 6 in school. A grade level score of 6 on a reading test is often interpreted as reflecting something about the degree of comprehension expected when the individual is reading material at grade 6 difficulty. However, the actual relationship between the grade level of reader ability and percent comprehension at different grade levels of difficulty is seldom, if ever, researched by developers of tests or developers of readability formulas.

The above list of problems does not exhaust all of the problems researchers in this area have encountered or will encounter. Rather than attempt to cover more of the problem, let us consider one of the more prominent solutions. The use of the cloze procedure is usually considered to provide a "better" solution to measuring readability (see Klare, Sinaiko, and Stolurow, 1972). Its advantage is that it involves an interaction between the reader and the material, (i.e., the higher the cloze score the more comprehensible the material is considered to be for the person completing the cloze blanks). In fact, the cloze score itself has been scaled by Bormuth (1968) to reflect percent comprehension (i.e., a 44% score on cloze is supposed to reflect 75% comprehension as reflected by a multiple-choice test). Bormuth (1967, 1968) has already discussed several problems with developing and using a criterion, such as 44%, but I want to focus upon what might be considered as the most prominent problem. Bormuth (1967) has argued that the 44% cut off, or any particular cut off score, is valid across ability levels. It seems reasonable that a cut off such as 44% would be valid for one level of ability usch as the fourth-, fifth-, and sixth-grade readers that have been used in much of this research (Bormuth, 1967; Bormuth, 1968; Rankin & Culhane, 1969). However, it seems unreasonable that this same cut off percent would be valid for high school students as well as college students. It seems likely that college students could fill in 44% of the blanks in material out of their field of expertise but their comprehension of this material upon reading the material once would be extremely low (e.g., 25% or less). On the other hand, it also seems reasonable that grade 2 or 3 students could comprehend 75% or more of passages that they could not complete more than 30% of the cloze blanks. These inconsistencies seem reasonable because as reading ability increases, the redundancies of language and grammar should become more apparent to the reader. Thus, graduate students are likely to be able to fill in a high percentage of the blanks in material that is incomprehensible to them while elementary students are likely to be less able to fill in blanks that are comprehensible to them simply because they are not able to take advantage of the redundancies of the language patterns. The cut off criterion for 75% comprehension probably increases as the grade level ability of the reader increases. In short, it is being argued here that the cloze cut off solution offers more than it can possibly deliver. Certainly, more research is needed on the interrelationships among cloze completion percentages, material difficulty as measured by readability formulas, individual ability as measured by standardized tests, and some independent measure of percent comprehension. At least, more research is needed to support the arguments of those who contend that the cloze cut off criteria are valid across all ability levels as indcators of percent comprehension.

In an effort to provide a better solution to some of the above discussed problems, Carver (in press, b) has developed a measure of reading ability which has been anchored directly to a measure of material difficulty. The measure of material difficulty is called the Rauding Scale of Prose Difficulty (Carver, 1974). The measure of individual ability is the National Reading Standards. A grade 6 score on the test, for example, can be interpreted as indicating that the individual should average 75% comprehension on passages at grade 6 difficulty according to the Rauding Scale. Empirical data (Carver, in press, a) indicate that: (a) when the difficulty of the material is three grade levels above the ability of the individual, then the individual probably has at least 75% comprehension of only 25% of the passages at this level, and (b) when the difficulty of the material is three grade levels below the ability of the individual, then the individual probably has at least 75% comprehension of 75% of the passages at this level. This approach and this research seems to lend more precision to the prediction of comprehension, given grade levels of material difficulty and grade levels of individual ability.

The results of an unpublished research study using the above measures of material difficulty and individual ability provide background for the earlier contention that the 44% cloze cut off is probably not valid across ability levels. The cloze technique was not used in the research but a highly similar technique was used, the reading input task; this task is similar to cloze and has been referred to as "second generation cloze" (Carver, 1975). Passages in second generation cloze at five levels of difficulty were administered to students at five levels of ability. Students at Grade 14 ability reading Grade 14 materials completed 96% of the second generation cloze task correctly. Students at Grade 11 ability reading Grade 11 materials completed 85% of the second generation cloze task correctly. The corresponding figures for grades 8, 5, and 2 were 83%, 77%, and 20%, respectively. The cloze scores consistently dropped in percent correct as the match between ability level and difficulty level dropped. However, the average percent comprehension was constant at 75% in all five of these situations. These data indicate that it would not be valid to suggest a constant cut off score for reflecting percent comprehension. Another way to articulate this point is to say that the relationship between percent correct on this reading task and percent comprehension is not the same at different levels of ability. It would seem to be an unlikely piece of luck for a similar task such as regular cloze to have exactly the same relationship across different levels of difficulty, (i.e., it seems unlikely that 44% cloze completions reflects exactly 75% comprehension across all levels of reading ability).

There are many situations where it would be advantageous to be able to predict what percent of a passage an individual would comprehend upon reading it once, given a measure of individual ability and material difficulty. This article has reviewed some of the problems involved in conducting this research and has reviewed some of the solutions that have been offered. The development of a valid measurement procedure which would allow a relatively accurate prediction of percent comprehension at all levels of difficulty and ability would seem to be worthy of attention from reading researchers.

References

Bormuth, J. R. Comparable cloze and multiple-choice comprehension test scores. *Journal of Reading*, 1967, *10*, 291–299.

Bormuth, J. R. Cloze test readability: criterion reference scores. *Journal of Educational Measurement* 1968, *5*, 189–196.

Carver, R. P. *Manual for the Rauding Scale Qualification Test*. Kansas City, Mo.: Revrac Publications, 1974.

Carver, R. P. Revised procedures for developing reading-input materials and reading-storage tests. *Journal of Reading Behavior*, 1975, *2*, 155–172.

Carver, R. P. Further research on the reading-storage test as a measure of gain during reading. *Journal of Reading Behavior*, in press. (a)

Carver, R. P. National Reading Standards. Kansas City, Mo.: Revrac Publications, in press. (b)

Cline, T. A. Readability of community college textbooks. *Journal of Reading*, 1972, *16*, 33–37.

Klare, G. R., Sinaiko, H. W., & Stolurow, L. M. The cloze procedure: A convenient readability test for training materials and translations. *International Review of Applied Psychology*, 1972, *21*, 77–106.

Rankin, E. F. & Culhane, J. W. Comparable cloze and multiple-choice comprehension test scores. *Journal of Reading*, 1969, *13*, 193–198.

ANALYTIC AND SYNTHETIC PROCESSES IN CONCEPT LEARNING AND THEIR RELATIONSHIP TO READING COMPREHENSION

JOCELYN CHEN
The University of Connecticut

RICHARD H. BLOOMER
The University of Connecticut

Concepts have long been associated with learning theory. The pioneer studies of Hull (1920) and Smoke (1932) demonstrated that subjects could increase their memory and discrimination ability if they saw concepts behind the stimuli presented. Since then, researchers in the area of reading have found that one's ability to organize and synthesize written material into meaningful concepts may be related to reading comprehension. In fact, several dimensions of reading behavior have been linked with concept development such as its being the difference between good and poor readers (Jan-Tausch, 1961; Kress, 1956), is more highly correlated with reading achievement than intelligence (Braun, 1963), is reflected in vocabulary knowledge (Chase, 1961; Langer, 1967), and is predictive of reading readiness in preschool children (Beller, 1970).

Though the findings of previous research clearly favor a relationship between conceptual ability and reading comprehension, two major weaknesses seem apparent in the studies found. First, interpretation of the results of the studies is limited by the level of concept ability actually measured. Concept ability was typically assessed by simple sorting tasks dealing with such concrete concepts as color, shape, and size. When verbal stimuli were presented, they usually were prepositions (e.g., over and under), comparatives (e.g., taller than or shorter than), or superlatives (e.g., biggest and smallest). Children's literature is hardly limited to the variety of concepts tested here and is more likely to include more complex concepts. Yet, it was inferred by the researchers that since good readers performed well on such concept tasks as those mentioned, concept ability was indicative of reading comprehension. One does not necessarily imply the other and it may well be that those who are good readers are also adept at many other tasks with simple concept learning possibly the lesser of those abilities. Second, few experimenters have concerned themselves with the underlying processes in concept learning, the study of which may suggest how concepts are formed and their subsequent effect on comprehension.

Two processes are of particular interest here since they both are fundamental to concept acquisition and development (Nelson, 1974) as well as instrumental to the understanding of English literature (Henry, 1974). Analysis, similar to concept attainment, is a deductive process requiring the learner to first discriminate a salient characteristic of a stimulus and then to generalize across a complex group of stimuli. Synthesis, similar to concept formation, is instead an inductive process relying not only upon discriminations made to an aspect of a stimulus set, but is a combinatorial process requiring responses to both similar and different dimensions of the stimuli in order to generate a higher-order or superordinate concept. Thus, it is concepts that are the product of synthesis while attributes are the result of analysis. Everyday examples of these processes range from such activities as identifying the theme of a story (analysis) or writing a sentence of one's own (synthesis) to a young child calling a cow a "doggie" (synthesis) or learning that a ball is something that can be thrown, rolled, held, bounced, picked up, and fit into a cup (analysis).

On the basis of previous investigations, then, the present study was designed to evaluate the relationship of analytic and synthetic processes in concept learning to reading comprehension using tasks involving verbal stimuli more closely related to that which would be found in children's reading material.

Method

Subjects

A group of 330 3rd-, 4th-, 5th-, and 6th-grade males and females of average ability from a white, lower middle-class community in southeastern Connecticut served as the subjects for this study.

Procedure

To determine each subject's analytic and synthetic ability, both the Analytic and Synthetic Concept Learning Tasks taken from the *Tests of Learning Processes* (Bloomer, 1975) were administered to groups of subjects in a single setting.

The first was an Analytic Concept Task in which subjects were presented with a set of things that go together. Then, given a second set of items, they had to answer yes or no as to whether each of the items went with the first set. Finally, Ss had to identify the concept being represented by the nonsense CVC given. This was repeated 13 times with 13 different concepts.

In the Synthetic Concept Task, Ss were presented with a set of nonsense CVCs paired with definitions. Ss were asked to remember each of these paired-associations and then synthesize only those nonsense CVCs whose definitions pertain to the concept mentioned to form a new name for the concept. Sixteen different sets ranging from 2 to 5 parts for the concept were included.

Both tasks were presented orally and required a written response. All nouns and adjectives used in the tasks were selected from among the 1,000 most frequent words of *The Teacher's Word Book of 30,000 Words* (1944), and were of moderate sonic affect and concreteness/abstractness. Application of the Spearman-Brown split-half reliability formula indicated reliability coefficients of .69 and .84 for the analytic and synthetic concept tasks respectively.

The number of correct responses for each of the two concept measures was recorded along with each subject's age, sex, IQ score from the most recent testing of the Otis-Lennon Mental Ability Test (Form J), reading comprehension and vocabulary scores from the SRA Achievement Tests (Blue Form C), and letter and word cloze scores from the Reading Skills Diagnostic Test (Bloomer, 1967).

Results

Mean performance across grade levels on the synthetic concept task showed inconsistencies with grade four performing better than grades 3, 5, and 6. By analysis of variance, these differences were found to be significant $\mathbf{F}\ (3,330) = 4.59$, $\mathbf{p} < .01$, but those between males and females were not.

Similar results were found on the Analytic Concept Task with small, but significant differences across grades with grade four exceeding the others $\mathbf{F}\ (3,330) = 7.29$, $\mathbf{p} < .01$. Though no significant differences were found between the sexes, there appeared to be a significant interaction between grade and sex $\mathbf{F}\ (3,330) = 3.54$, $\mathbf{p} < .05$.

To determine the relationship among the variables, a standard Pearson **r** formula was applied. The resulting correlation matrix for the entire sample showed significantly low positive correlations ($\mathbf{r}$=.31 to .39; $p < .01$) for both concept tasks and the variables IQ, reading comprehension, vocabulary, letter cloze, and word cloze. When males and females were analyzed spearately, quite a different picture appeared as to the nature of the relationships among the variables. Analytic and synthetic concept abilities among females were more closely related to such variables as IQ, reading comprehension, vocabulary, letter cloze, and word cloze. In fact, the intercorrelations among these last five variables were higher and more consistent for females than males.

Finally, the results of a stepwise multiple regression analysis suggested that in the analytic concept task, letter cloze accounted for most of the variance for both sexes (11% for males and 30% for females) and, likewise, in the synthetic task for males (12%) though for females, variance was mainly attributed to vocabulary (19%).

Discussion

While the study did not show analytic and synthetic concept ability to be strongly related to reading comprehension, the results did reveal some interesting differences between males and females that may explain why females seem to do better than males in the elementary grades. The finding that the variables IQ, reading comprehension, vocabulary, cloze, and concept ability were all intercorrelated to the same degree for females suggests that females are able to transfer their skills from one verbal task to a new, but similar one, thereby supporting a general intelligence theory. Boys, on the otherhand, showed expected correlations for the variables reading comprehension, vocabulary, and cloze with IQ which is partially dependent on reading ability. On the concept tasks, however, there was a significant decrease in the correlations with these other factors. Thus, there seems to be little transfer of proficiency in one verbal task to others of a similar nature in boys. The flexibility displayed by girls in cognitive functioning may account for the differential reading abilities found in favor of them as reported by previous research in reading.

The findings presented here have implications for methodological considerations in the teaching of reading or of any other type of learning. Most existing curricular programs rely heavily on the learner's ability to generalize across skills which, according to the results presented here, would hinder progress among males. Instead, males would learn more efficiently under a program offering a criterion-referenced or mastery model within skills.

Further study is needed to explore other possible differences between males and females in associative learning ability, particularly as it pertains to abstract or complex stimuli.

References

Beller, E.K. The concept of readiness and several applications. *The Reading Teacher*, 1970, *23*, 727-737.

Bloomer, R.H. *The reading skills diagnostic test.* New York: Brador Publications, 1967.

__________ *Tests of learning processes.* Mimeo. The University of Connecticut, Storrs, 1975.

Braun, J.S. Relation between concept formation ability and reading achievement at three developmental levels. *Child Development*, 1963, *34*, 675-682.

Chase, C.I. An application of levels of concept formation to measurement of vocabulary. *Journal of Educational Research*, 1961, *55*, 75-78.

Henry, G.H. *Teaching reading as concept development: Emphasis on affective thinking.* Delaware: International Reading Association, 1974.

Hull, C.L. Quantitative aspects of the evolution of concepts; An experimental study. *Psychological Monographs,* 1920, *28,* 123-154.

Jan-Tausch, J. Concrete thinking as a factor in reading retardation. *Dissertation Abstracts,* 1961, *21,* 2984-2985.

Kress, R.A. Jr. An investigation of the relationship between concept formation and achievement in reading. *Dissertation Abstracts,* 1956, *16,* 573-574.

Langer, J.H. Vocabulary and concept development. *Journal of Reading,* 1967, *10,* 448-456.

Nelson, K. Concept, word, and sentences: Interrelations in acquisition and development. *Psycnological Review,* 1974, *81,* 267-285.

Smoke, K.L. An objective study of concept formation. *Psychological Monographs,* 1932, *42,* 46-157.

Thorndike, E.L. & Lorge, I. *The teacher's word book of 30,000 words.* New York: Bureau of Publications, Teachers College, 1944.

THE INTERACTIVE EFFECTS OF VERBAL COMPREHENSION, NONVERBAL PERFORMANCE, REVIEW QUESTION TYPE, AND REVIEW QUESTION PLACEMENT ON READING COMPREHENSION

RANDALL A. SILVERSTON
University of Texas at Arlington

R. JON LEFFINGWELL
University of Texas at Arlington

DON M. BEACH
University of Texas at Arlington

JAMES BENSON
University of Texas at Arlington

ABSTRACT

Statistical interactions between measures of certain cognitive abilities, placement of review questions in text, and review question type (higher order or lower order) were posited as being predictive of reading comprehension abilities. The Henmon-Nelson Intelligence Test was administered to 76 undergraduate students so as to derive verbal, quantitative, and total scores. Students were then assigned to 1 of 5 textual treatment conditions (based on review question type and review question location variables) through a stratified sampling technique based upon their total scores on the Henmon-Nelson test.

A comprehension test was designed so as to identify intentional and incidental learning. All participants were administered the comprehension instrument immediately after the textual presentations. As a motivational control, subjects were informed that they were to be paid 15 cents for each correct answer on the comprehension test.

None of the hypothesized interactions were supported. Post hoc analyses revealed: (a) The sample may have been regressed due to the motivational contingencies evident in the study. (b) A significant interaction was found to exist between quantitative scores on the Henmon-Nelson test and lower order and higher order questions to the extent that subjects who scored lower on the quantitative measure appeared to do better on the comprehension instrument if they received lower order questions in text. Conversely, subjects who scored higher on the quantitative measure appeared to do better on the comprehension instrument if they received higher order questions.

THE CONTRIBUTION OF TESTS OF SEMANTIC AND SYNTACTIC CONTEXT CLUES TO THE PREDICTION OF TWO READING COMPREHENSION MEASURES

JAMES W. CUNNINGHAM
UNC–Chapel Hill

PATRICIA M. CUNNINGHAM
UNC–Chapel Hill

Valid scores representing subjects' reading comprehension are necessary for reading diagnosis, evaluation, and research. Several procedures for obtaining comprehension scores are currently in use, among them cloze tests of various types, tests utlizing comprehension questions in various formats, and retelling outlines (Goodman & Burke, 1972). Because each of these procedures probably represented a kind of "lateral thinking" (de Bono, 1970) when it first appeared, these types of tests seem to have had no rigorous validation as part of their development but instead began with a kind of **face validity**. Consequently, much debate and research has occurred attempting to explain or defend the **validity** or lack of it for particular comprehension scores gained by a particular procedure. Continuing debate and research offer no promise of a resolution, however, possibly because the procedures exist and are measuring something or some things. Unfortunately, present logical and empirical abilities preclude knowing exactly what, and in what proportions, these things are.

What is needed is a **decision rule** and validation designs which bring this rule to bear on comparisons between testing procedures. The effort to determine the quality and quantity of absolute **validity** possessed by a set of scores should continue, of course, but for practical purposes, means to determine which of several testing procedures is more **valid** are needed.

This study is an example of a **criteria-related** validation design using the following **decision rule: If sets of scores (criteria are valid measures of components of a process, the more valid global measure of that process will be that global measure correlating more highly with the criteria**.

Problem

In recent years, information-processing theory has had an increasing influence on the study of the reading act. Researchers find a growing body of theoretical and experimental literature dealing with the intersection of information processing and reading. Within this literature, there appears to be wide agreement that reading involves cue sampling (Walker, 1974-1975), although there is some problem with determining which specific cueing systems are involved (Siler, 1973-1974). Goodman (1965) defined three major sources of cues to be grapho-phonic, syntactic, and semantic. Others such as Weber (1968) and Smith (1971) have lent support to the uniqueness and importance of syntactic and semantic constraints in the reading process.

Assuming the crucial place of syntactic and semantic cueing in processing information through reading, any test of reading ability ought to measure, at least in part, the competence of the reader to utilize syntactic and semantic cues. Although miscue analysis (Cunningham, 1975, Goodman, 1969; Goodman and Burke, 1972) is a diagnostic and evaluative procedure which attempts to measure competence with the three cueing systems suggested by Goodman, traditional standardized and informal measures do not obviously evaluate ability to respond to syntactic and semantic cues even if such ability is measured indirectly by any test which requires reading of contextual material to obtain meaning.

At some point it may be possible to determine the relative emphasis which test makers should place on syntactic and semantic constraints. At this point there is some question as to which, if either, is more constraining and to what degree. The purpose of this study was to determine if a cloze test of a kind often used by classroom teachers and the reading comprehension subtest of the Gates-MacGinite Reading Tests (1965) were measuring syntactic and semantic abilities equally or if they differed in their measuring of these abilities.

Method

Subjects

All 120 fifth and sixth graders in an elementary school in Southeastern Ohio were the subjects for the study. Eleven students were absent at least one of the 4 days of testing and were not included due to incomplete data on them. The remaining 109 students became the population for the study.

Tests

Cunningham Test of Ability to Use Semantic Context Clues. The investigators selected 15 words (5 adjectives, 5 nouns, and 5 verbs) with which, in their judgment, fifth and sixth graders would probably not be familiar. Each of these 15 content words was introduced in a statement of one or more sentences which gave semantic context clues to a meaning for that word. The test itself consisted of 15, 1-sentence statements, each of which had a 1-word blank of uniform length. In each case, the word which had been deleted was one of the 15 words introduced in context. Each item of the test was followed by the 5 words of that form-class appropriate to the item. The effect of memory was minimized in that subjects were instructed to refer to the statements introducing the words as they took the test. Of course, the ability to use syntactic constraints was required in order to read the statements introducing the words and the items themselves. Siler (1973-1974) concluded that semantic constraints were dependent on syntactic constraints. Subjects, however, were unable to use syntactic information to improve their score on the items of the test. Because all distractors were of the same form-class as the correct response, the subject was given the correct form-class for the answer whether or not he could have utilized synatctic constraints to predict the appropriate form-class. And the subject had to select between the response-choices based on semantic constraints alone.

Cunningham Test of Ability to Use Syntactic Context Clues. A brief, 40-word passage was used from which most meaning had been deleted by replacing content words with nonsense words having endings indicative of form-class. This largely meaningless passage was followed by 10 multiple-choice questions, most of which were abbreviated **wh**-, transformation questions (Finn, 1973). Response-choices were all words from the passage after the meaning-removal process.

Cloze Test. A passage was systematically selected from the middle of a story in a supplementary basal the subjects had never seen. The first sentence was left intact. Beginning with the second sentence, every fifth word was deleted and replaced by a blank of uniform length until 25 words had been deleted. The last sentence was left intact. By this random procedure, 6 nouns, 1 pronoun, 6 adjectives, 3 adverbs, 4 verbs, 3 conjunctions, and 2 prepositions were deleted. The readability of the passage as reported by the teacher's manual was fourth-reader level.

Gates-MacGinite Reading Tests, Comprehension Test. Survey D, Form 2 of the Gates-MacGinite was used.

Procedure

The four tests were administered to groups of children on 4 different days over a 2-week period. The actual testing was handled by the principal of the school and the investigators. The Gates-MacGinite was timed according to the test manual. Students were given all the time they needed for the other three tests. Particular care was taken that students did not help and were not helped by each other.

Data Analysis

To determine if a cloze test and the comprehension section of the

Gates-MacGinite were measuring syntactic and semantic context-clue abilities equally, two multiple regression analyses were computed among raw scores from the four tests. In both analyses, the syntactic and semantic raw scores were entered as predictors: in the first analysis, the Gates-MacGinite raw score was the criterion, in the second analysis, the cloze raw score was the criterion. The resulting two multiple correlation coefficients were tested for the significance of their difference by a dependent test for the difference between correlations obtained from the same sample. If this difference were significant, the results were interpreted by an examination of the semipartial correlation coefficients generated by the multiple regression analyses. An alpha-level of .01 was established for all significance testing.

Results

Both multiple correlation coefficients were significant. The correlation between the syntactic-semantic composite and the Gates-MacGinite Comprehension scores was .780; the correlation between the syntactic-semantic composite and the cloze scores was .651.

After it was determined that both Multiple R's were significant, a two-tailed test of the difference between dependent correlations was computed and found significant $t\ (106) = 2.665,\ p < .01$. The correlation between the syntactic-semantic composite and the G-M Comprehension was significantly greater than the correlation between the syntactic composite and the cloze.

After these multiple correlations were found to differ, selected semipartial correlation coefficients were extracted from the multiple regression analyses.

Table 1
Selected Semipartial Correlations
Among Syntactic, Semantic, G-M, and Cloze Scores

Semipartial r	Variance Accounted For
$r_{\text{g-m (syn. sem)}} = .386$	14.887%
$r_{\text{g-m(sem. syn)}} = .297$	8.839%
$r_{\text{clz(syn. sem)}} = .092^{*}$	0.849%
$r_{\text{clz(sem. syn)}} = .452$	20.439%

N = 109

* = Not Significant

Disucssion

Limitations of the Study

1. The Semantic test is not completely free of syntactic constraints.
2. The Syntactic test is not completely free of semantic constraints.
3. The cloze test used may not have been representative of all cloze tests.
4. The population for the study was from one school in one part of the country.
5. The number of subjects presented problems in interpreting semipartial correlation coefficients.
6. Multiple significance-testing lowers the true alpha-level.

Interpretation of the Results

Within the limitations of the study, certain cautious conclusions may be reached. Given the **decision rule**, the Gates-MacGinite (G-M) reading comprehension subtest (Survey D, Form 2) appears to be a better measure of reading comprehension than the specific cloze test used, for this population of 109 fifth and sixth graders. It would appear that some of the G-M's superiority was due to the lack of syntactic clue use required by the cloze test.

The design has resulted in a decision concerning the more **valid** of two testing procedures for a specific population. Obviously, more confidence could be placed in the decision if the study had fewer limitations, but this lack of confidence is not a weakness in the **decision rule** but in the evidence on which the decision was based.

References

Cunningham, J.W. *A Simplified Miscue Analysis for Classroom Teachers.* Unpublished mimeograph, University of North Carolina, 1975.

de Bono, E. Lateral and vertical thinking. *Today's Education,* 1969, *58* (8), 20-24.

Finn, P.J. An algorithm for deriving operationally defined comprehension questions from written texts. Unpublished doctoral dissertation, University of Chicago, 1973.

Gates, A.I., & MacGinite, W.H. *Gates-MacGinite Reading Tests* (Survey D). New York: Teachers College Press, 1965.

Goodman, K.S. Analysis of oral reading miscues: applied psycholinguistics. *Reading Research Quarterly,* 1969, *5,* 9-30.

Goodman, K.S. A linguistic study of cues and miscues in reading. *Elementary English Journal,* 1965, *42,* 39-44.

Goodman, Y.M., & Burke, C.L. *Reading miscue inventory.* New York: Macmillan Publishing Co., 1972.

Siler, E.R. The effects of syntactic and semantic constraints on the oral reading performance of second and fourth graders. *Reading Research Quarterly,* 1973-1974, *9,* 583-602.

Smith, F. *Understanding reading.* New York: Holt, Rinehart, and Winston, 1971.

Walker, L. A comparative study of selected reading and listening processes. *Reading Research Quarterly,* 1974-75, *10,* 255-257.

Weber, R. The study of oral reading errors: a survey of the literature. *Reading Research Quarterly,* 1968, *4,* 96-119.

EFFECT OF READER–GENERATED QUESTIONS ON LEARNING FROM PROSE

JANE M. MORSE
University College
Rutgers – The State University of New Jersey

The study investigated the effects of reader generation of questions while studying textbook-type material. An attempt was made to ascertain whether questions composed by the reader facilitated learning more than experimenter-generated ones or no questions at all. Position effects of the questions were investigated. Further, the effects of questions for each passage were examined. In the special case where readers generated questions both before and after reading, the effect of extending working time was analyzed.

The adjunct question paradigm was chosen for the investigation which calls for insertion of questions either before, after, or within a prose text. The effect of these interspersed questions is then examined by a posttest over the content of the passage.

The available literature is ambiguous about the effect of interspersed questions on learning. Although Rothkopf (1966), Rothkopf and Bisbicos (1967), Frase (1967, 1968a, 1968b), and Boker (1974) reported that both prereading questions and postreading questions facilitated learning (but to different degrees), Boyd (1973), Hiller (1974), Richmond (1974), and Washburne (1929) have reported findings to the contrary.

The differences in findings may be attributable to a number of variables: (a) Learning time was not controlled in some studies. (b) Review was permitted in some studies and not in others. (c) Some studies used only one passage while others varied the material. (d) There were differences in statistical procedures. (e) Directions given to subjects differed from one study to another.

In order to account for or control these several variables in this investigation, inspection time was controlled, review was permitted to make the task more realistic, passages on three different topics were used, Dunnett's test was used to compare the treatment groups to the control group, and subjects were given explicit directions as to the nature of the posttest questions.

Design of Study

The design was a 2 x 3 x 3 factorial analysis of covariance with repeated measure on the c factor with the combined verbal and comprehension scores on the Nelson-Denny Reading Test serving as the adjusting variable. The factors for this design were: (a) experimenter or reader generation of questions, (b) questions before, questions after, or questions before and after, and (c) passages (F, M, V).

Because of the difficulty in composing questions both before and after for the RBA (reader-generated before and after condition) group, in the allotted 7 minutes, a separate analysis was made substituting the scores of the group given 11 minutes to complete writing preview and review questions and reading.

Subjects

One hundred and eight college undergraduate students in an Introduction to Psychology course at a four-year college were randomly assigned to 1 of 9 groups.

Materials

The materials were three prose passages of approximately 600 words each written by the experimenter from information taken from three college-level textbooks. Each passage was constructed from an outline which had an introduction and three main points, with two or three subordinate points under each main item. The titles were *The Federal Reserve System*, *Music of the Baroque Period*, and *Viruses*.

In addition to the passages, there was a title page containing the title, the three subtitles, and a brief summary statement about the entire passage.

There were 10 multiple-choice questions constructed for each passage: one question generated from the introductory paragraph, one question for each of the main subtopics, and one question for each of the subordinate points.

The materials also included six experimenter-written questions for each passage which were used for those conditions requiring them. The three used for preview were generated by turning the subtitles into questions. The three used for review questions were generated from a subordinate point of each of the three main topics. These questions were all of an open-ended type. For example: What are some characteristics of Baroque music?

The Nelson-Denny Reading Test, Form D, (Brown, 1973) was used to assess general reading level for the subjects. The combined vocabulary and comprehension score from this test was used as covariate for the data analysis.

Directions

The eight different sets of directions gave the purpose for the study, the specific orders of tasks to be done, and the time limits. It was explicitly stated that there would be a multiple-choice test at the end of each passage so that the subjects would know exactly the nature of the posttest.

Results

Question Effects

A Dunnett's test (Winer, 1971) was used to compare the means of each treatment group with the means of the control group. There were no significant differences between the treatment groups and the control group when the means were adjusted by the covariate.

Generation, Position, and Passage Effects

The analysis of covariance used data only from the groups which actually utilized questions before, after, and before and after. The means of experimenter–and reader-generation conditions across passages were 8.3 and 7.7, respectively, $\mathbf{F}\ (1, 65) = 4.35$, $\mathbf{p} < .05$. Using formulas developed by Clark (1973), quasi-**F**s were calculated for source of generation with passages treated as random variables. The differences among generation conditions remained significant, $\mathbf{F}'\ (2.6, 1715) = 3.52$, $\mathbf{p} < .05$, even though analysis of passages separately yielded a significant difference only for the **Viruses** passage, $\mathbf{F}\ (1, 65) = 5.58$, $\mathbf{p} < .05$.

There were no significant differences in these analyses between groups who received or generated questions before, after, or before and after. There was a significant difference between groups which read experimenter-generated questions and those groups which generated their own.

Extended Time Effects

The RBA group with extended time had a higher mean (8.3) than the group which had insufficient time (7.4). An analysis of covariance with the scores of this extended time group substituted for the group with the seven-minute time was computed. There were no significant differences among any of the groups on any of the passages.

Correlation of Experimental Task and Reading Aptitude

The correlation coefficients ($\mathbf{r} = .59$, $\mathbf{p} < .001$) between scores on the experimental task and the Nelson-Denny Reading Test indicated that those who performed well on the Nelson-Denny Reading Test tended to score well on the passage tests.

Discussion

Question Effects

The results of the study indicate that reading textbooktype materials with no questions is just as effective as reading with either reader-generated or experimenter-generated questions. This must be qualified to pertain when review is allowed, when inserted questions are of an open-ended type dealing with main ideas, and when the posttest is a recognition-type test.

There are some possible explanations for the lack of significant effects in the study. It may be that the nature of the experimenter-generated questions asked in the study were reiterative; that is, they required recall of exact material in the passages; and, thus, were not as helpful as "application-type" questions might have been (Watts & Anderson, 1973). It is also possible that because of the change in form between inserted questions and posttest questions, the inserted questions failed to be facilitative.

The mean scores on the posttest were very close to a perfect score. The posttest ceiling may have been too low to allow appropriate spread in scores. However, there were only three subjects of the 108 who achieved a perfect score of 30; and the range was from 8 to 30.

The highly structured nature of the material may have aided recall for groups equally – there is some evidence that preknowledge of structure has a positive effect on learning (Ausubel, 1960).

The effects of review may also account for lack of differences since Hiller (1974) found that review was a factor that made his results different from previous studies.

It may be that training is necessary before self-questioning is a usable tool. A.E. Smith (1972), Donald (1967), and Borthwick (1975) had success in training students in questioning techniques.

Experimenter or Reader Generation and Passage Effects

Experimenter-generated questions were relevant questions; that is, they pointed to information that would be covered on the posttest. It might have been predicted that such questions would facilitate learning more than reader-generated ones. This did, indeed, seem to be the case since the Clark (1973) quasi-F remained significant across passages for generation conditions. However, the effect appeared conclusively for only one passage. For the *Viruses* passage, experimenter-generated questions seemed to be of greater benefit than reader-generated ones. Since the *Viruses* passage had the highest readability level, the interpretation might be made that experimenter-generated questions tend to aid more on more difficult material.

In the analysis of covariance, there were no significant differences among passages; and passages entered no significant interactions with other variables. Thus, the pattern of effects of passages supports the hypothesis of no difference due to passages.

Extended Time

The RBA group (Reader Generated Questions Before and After) had the lowest mean scores of all the groups when working time was held constant. This group had the most operations to perform in an equal amount of time and could not complete the composing of postquestions. This depressed their scores, presumably because they were occupied with composing questions rather than learning the material (Mackworth, 1966). The group with extended time completed the question-asking task, but their scores were not significantly different from any other group. This suggests they were also occupied with questioning and not with learning since it might otherwise be presumed that because this group had more exposure time to the materials, it would have the highest posttest scores of all (Nodine & Goss, 1969).

It also suggests that reader generation of questions both before and after reading is an inefficient way to study text-book-type materials for a multiple-choice test since the method takes longer than other strategies to achieve similar results.

A review of pertinent literature has shown that if students are provided with questions and answers which are to appear on a subsequent test, improvement in test scores might be anticipated. But without the same questions, students might be handicapped with questions because of having attention directed to irrelevant material. The lack of significant differences among groups in this study indicates that caution should be used in recommending that questions be utilized as study aids. Certainly the source of questions, nature, and structure of the materials, the type of posttest expected, the amount of study time available, and the nature and specificity of questions asked are all variables to be taken into account.

Conclusions

The study found no facilitative effects for experimenter or reader generation of questions, for questions before or after, or .for questions before **and** after. Specifically, it was found that:

– Groups who utilized questions, either experimenter or reader generated, did not differ significantly from the control groups which had no questions.

– The position of the study questions – before, after, or before and after – did not affect posttest scores.

– Experimenter-generated questions proved to be more facilitative than reader-generated ones.

– There were no differences attributable to passages alone.

– There were no significant differences among groups when the scores of the group which had extended time for the generation of questions both before and after reading were substituted for the group which had more limited time.

References

Ausubel, D.P. The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology*, 1970, *51*, 267-272.

Boker, J.R. Immediate and delayed retention effects of interspersing questions in written instructional passages. *Journal of Educationcl Psychology*, 1974, *66*, 96-98.

Borthwick, P.B. *The effects of question strategy training upon the information gained from reading science-related prose selections with fifth grade students.* Unpublished paper presented at the National Reading Conference, Kansas City, December 1974.

Boyd, W.M. Repeating questions in prose learning. *Journal of Educational Psychology*, 1973, *64*, 31-38.

Brown, J.I., Nelson, M.J., & Denny, E.C. *The Nelson-Denny reading tests*, Forms C & D. Boston: Houghton Mifflin Co., 1973.

Bull, S.G., & Dizney, H.F. Epistemic-curiosity-arousing prequestions: Their effect on long-term retention. *Journal of Educational Psychology*, 1973, *65*, 45-49.

Carver, R.P. A critical review of mathemagenic behaviors and the effect of questions upon the retention of prose materials. *Journal of Reading Behavior*, 1972, *4*, 93-119.

Clark, H.H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 335-359.

Coleman, E.B. Generalizing to a language population. *Psychological Reports*, 1964, *14*, 219-226.

Dale, E., & Chall, J. A formula for predicting readability instructions. *Educational Research Bulletin*, Ohio State University, 1948, *17*, 37-54.

Donald, M. The SQ3R method in grade seven. *Journal of Reading*, 1967, *11*, 33-35; 43.

Frase, L.T. Learning from prose material: Length of passage, knowledge of results, and position of questions. *Journal of Educational Psychology*, 1967, *58*, 266-272.

Frase, L.T. Effect of question location, pacing, and mode upon retention of prose material. *Journal of Educational Psychology*, 1968, *59*, 244-249. (a)

Frase, L.T. Questions as aids to learning: Some research and theory. *American Educational Research Journal*, 1968, *5*, 319-332. (b)

Fry, E. Graph for estimating readability. In *Reading instruction for classroom and clinic.* New York: McGraw-Hill, 1972. P. 232.

Hiller, J.H. Learning from prose text: Effects of readability, level, inserted question difficulty and individual differences. *Journal of Educational Psychology*, 1974, *66*, 202-211.

Ladas, H. The mathemagenic effects of factual review questions on the learning of incidental information: A critical review. *Review of Educational Research*, 1973, *43*, 71-82.

Mackworth, J.F. Perceptual coding as a factor in short-term memory. *Canadian Journal of Psychology*, 1966, *20*, 18-33.

Natkin, G., & Stahler, E. The effects of adjunct questions on short and long-term recall of prose materials. *American Educational Research Journal*, 1969, *6*, 425-432.

Nodine, C.F., & Goss, A.E. Temporal parameters in paired-associates learning. *Psychonomic Monograph Supplements*, 1969, *3*, 1-17.

Richmond, M.S. *The relationship of the uniqueness of prose passages to the effect of question placement and question relevance on the acquisition and retention of information.* Unpublished paper presented at the National Reading Conference, Kansas City, December 1974.

Rothkopf, E.Z. Learning from written instructive materials: An exploration of the control of inspection behavior by testlike events. *American Educational Research Journal*, 1966, *3*, 241-249.

Rothkopf, E.Z., & Bisbicos, E.E. Selective facilitative effects of interspersed questions on learning from written materials. *Journal of Educational Psychology*, 1967, *58*, 56-61.

Washburne, J.N. The use of questions in social science material. *Journal of Educational Psychology*, 1929, *5*, 321-359.

Winer, B.J. *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill Book Co., 1971.

TESTING COMPREHENSION OF THE CENTRAL THOUGHT: SELECTING VERSUS GENERATING MAIN IDEA

GINGER VAN BLARICOM
University of Wisconsin-Madison

SANDRA WHITE
University of Wisconsin–Madison

Introduction

Comprehending the central thought, or main idea, of a passage is a comprehension skill commonly applied at all grade levels. In almost every classroom, children are taught to identify the main idea of passages via activities that require them to select, state, or write a topic, title, or main idea statement. Most often, the ultimate objective of these activities is to develop the ability to generate the main idea while reading independently (i.e., "going away" from printed material with a general idea of what the selection is about). Standardized and criterion-referenced tests often measure the ability to comprehend the main idea by asking the child to **select** from 3-5 choices, the topic, title, or main idea statement that most accurately reflects the central thought of a passage. Several critics question the validity of the multiple-choice test format for determining the ability to generate a main idea while reading independently. One such critic, Farr (1969) indicated that most reading comprehension tests have questionable construct validity; for, he says, what is measured most often, when children select rather than generate responses, is test-taking performance rather than reading ability or skill. Osenburg (1962), another critic of the multiple-choice test format, examined the variations between objective test scores and subjective evaluations of high school and college students. In his study, 400 subjects responded to a single test item written in two formats: first, the subjects were asked to generate a topic of a poem; then, each subject was asked to select an appropriate topic for the same selection. Forty-three percent of the Ss who responded correctly to the question in the selection format were not able to generate an appropriate response. Osenburg concluded that the multiple-choice objective test may not be measuring reading at all, but measuring instead the child's ability to distinguish terms, details, and facts from the passage from those with which the child is not familiar.

It would seem logical to conclude from the above that **selecting** a main idea statement and **generating** a main idea statement are different tasks, at least when the selection test requires little more than discriminating familiar from unfamiliar terms. Whether the two tasks differ when the selection test asks the subject to distinguish passage details from passage main idea has not been determined, however. This paper is a description of two experiments that investigated that issue. The purpose of the first experiment was to determine whether there would be any difference between selecting a main idea statement (on an objective test that required the subject to discriminate passage details from main idea) and generating a main idea statement. The second experiment was done subsequent to the first to corrobate the findings of Experiment 1, and to examine whether organizing the passages explicitly with topic sentences had an effect on selection/generation response differences.

Experiment 1

Purpose

As stated in the introduction, the purpose of Experiment 1 was to determine whether there would be any difference between selecting and generating a main idea statement. The following question was asked in this experiment:

> Will performance on a main idea selection test differ significantly from performance on a main idea generation test?

Subjects

One hundred children served as subjects for Experiment 1. The Ss were selected on a random basis from a population of 175 fifth-grade children in a semi-rural Wisconsin public school.

Materials

Twenty paragraphs were developed for this study. Each paragraph ranged in length from 60 to 80 words, and was written at a readability level of between 4.3 and 4.5 according to the Dale-Chall readability formula. None of the paragraphs had topic sentences. Two main idea questions were written for each of the 20 passages. One question asked for main idea selection; the other required main idea generation. For the selection items, the S was asked to select the 1 sentence of 4 that most accurately stated what each passage was all about. The incorrect answer choices for each selection item were passage details. For each generation item, the child was asked to write, independently, a complete sentence that would best tell what each passage was all about.

Testing Procedure

Two test forms were developed with the 20 passages. Each form included 10 passages with selection questions followed by 10 passages with generation questions. The forms were identical in content, differing only in that the passages that required selection responses on one form asked for generation responses on the other, and vice versa. The Ss were assigned randomly to test form, with 50 Ss responding to each. The tests were administered by classroom teachers, and were not timed. The Ss worked independently, following example items. Specific directions were given for writing complete sentences on the generation test. Ss were helped with pronunciation of unknown words, when necessary.

Scoring of Generation Responses

Prior to scoring the generation responses, the minimum response acceptable for each of the 20 passages was determined. A minimum response was defined as a complete sentence including the basic elements or concepts of the passage. The basic elements, or concepts, for each passage were determined a priori. All generation responses were scored independently by the two experimenters as either right or wrong. The two evaluations were compared, and when disagreement occurred, a third evaluation was made by an additional judge. Minimum responses defined as complete sentences including the basic elements or concepts of the passage were referred to as Generation 1 responses in the sections that follow.

In the process of scoring the generation responses according to the predetermined criteria, it became obvious that many Ss were able to generate the major elements or concepts of a passage, but did not generate complete sentences. So as not to penalize those Ss who did not write a complete sentence, the responses were evaluated a second time. In the second evaluation, a minimum response was defined as any response, whether a complete sentence or not, that included the basic elements or concepts of the passage. The same procedure for scoring was followed as in the first evaluation. In the sections that follow, these responses were referred to as Generation 2.

Results

Analysis of Variance. The design for Experiment 1 was a randomized-block with repeated measures on one factor. The factor studied was item type (i.e., selection, generation). Maximum score for each cell was 10.0. A look at the mean responses for each item type revealed that Selection scores (9.15) were higher than Generation 1 and 2 scores (3.36 and 6.09, respectively); and that Generation 2 scores were greater than Generation 1. The analysis of variance for Selection and Generation 1 and selection and Generation 2 responses indicated selection and generation differences were significant, $\mathbf{F}(1,99) = 412.45$, $\mathbf{p} < .05$ and $\mathbf{F}(1,99) = 172.42$, $\mathbf{p} < .05$, respectively for Selection and Generation 1 and Selection and Generation 2. Neither the subjects variable nor the interaction was significant. Test form was not a variable studied as the forms were considered equivalent.

Mastery/NonMastery Data. Criterion-referenced tests often report performance in terms of test mastery or nonmastery. Although the definition of mastery often varies from test to test, many define mastery as responding correctly to 80% or more of the items on a particular test. For descriptive purposes, the Ss' scores for selection and generation items in this experiment were analyzed for mastery. Mastery in this study was defined as 80% of the items correct (i.e., correct response to 8 of the 10 selection or generation items). A comparison of mastery levels given mastery of a particular subtest identified 89 of the 100 Ss as masters of the Selection test. Of these 89 Ss, 12 (14%) were also masters of the Generation 1 subtest, and 32 (36%) were masters of the Generation 2 test. Only 12 of the 100 Ss were masters of the Generation 1 subtest, and 33 were masters of the Generation 2 test. Interestingly, 100% of the Generation 1 and 97% of the Generation 2 masters were also Selection masters.

Experiment 2

Purpose

Substantial differences were found between selection and generation responses in Experiment 1. For each of the passages in Experiment 1, the Ss were asked to select or generate the main idea implied by the passage details. None of the passages in Experiment 1 had topic sentences.

Evidence from an earlier study (Van Blaricom & White, 1975) indicated that main idea comprehension is greater on passages with organizers or topic sentences than on passages without. In this second experiment, the experimenters were interested in determining the effect of passage organization (whether the passages were organized explicitly for the Ss via topic sentences) on selection/generation response differences. The following questions were asked in Experiment 2:

> Will performance on a main idea selection test differ significantly from performance on a main idea generation test when the passages are organized explicitly (i.e., include topic sentences)?
>
> Will performance on a main idea selection test differ significantly from performance on a main idea generation test when the passages are not organized explicitly (i.e., do not include topic sentences)?

Subjects

One hundred and twenty children were the subjects for Experiment 2. The Ss were selected at random from over 200 fifth-grade children in two public schools in Wisconsin. Sixty Ss were chosen from each school. Neither school participated in Experiment 1.

Materials

Forty paragraphs were developed for this experiment. Twenty of the paragraphs were those used for Experiment 1. The other 20 were identical except topic sentences had been added either at the beginning, in the middle, or at the end of the various passages. The selection and generation questions for each of the 40 passages were identical to the questions given in Experiment 1.

Testing Procedure

Four test forms were developed with the 40 passages. Each form included 10 passages with selection questions followed by 10 passages with generation items. Forms 1 and 2 were identical in content, differing only in that the passages that required selection responses on one form asked for generation responses on the other, and vice versa. Forms 3 and 4 differed from 1 and 2 respectively only in that the passages had topic sentences. The Ss were randomly assigned to 1 of the 4 test forms, with 30 Ss responding to each. The procedure followed for administering and scoring the tests was identical to that given in the first experiment.

Results

Analysis of variance. The design for Experiment 2 was a split-plot factorial with repeated measures on the within factor. The between factor was paragraph organization – no organizer (i.e., topic sentence) organizer, and the within factor was item type – selection, generation. The maximum score for each cell was 10.0 From an examination of the mean scores, it was evident that selection scores were greater than generation scores, which was a finding that was consistent with that of Experiment 1. Across paragraph organization, the mean response for Selection items was 9.0, while for Generation 1 and 2 items it was 4.42 and 6.93, respectively. Within paragraph organization, the means were 9.17, 5.02, 7.42 for passages with organizers and 8.83, 3.93, and 6.43 for passages without organizers. The analysis of variance for Selection and Generation 1 responses and Selection and Generation 2 indicated that all main effects for both analyses were significant at the .05 level. The Selection and Generation 1, and the Selection and Generation 2 differences were significant with $F\ (1,118) = 274.288$, $p < .05$ and $F\ (1,118) = 131.496$, $p < .05$ respectively. The organization variable was significant for both analyses as well. No significant interaction between factors was found.

Mastery/nonmastery data. The Ss' scores for selection and generation items were analyzed for mastery in this experiment as they were for Experiment 1. Following the procedure described in Experiment 1, mastery was defined as 8 of the 10 responses correct for each subtest. Using this criterion, 85.8%, 23.3% and 47.2% of the Ss were identified as masters of Selection, Generation 1, and Generation 2 items respectively. Looking within paragraph organization, the percent mastery is greater for all question types on passages with as opposed to passages without organizers. The number of Generation 1 masters on passages with organizers is nearly double the number on passages without. Although the differences were not as substantial, 56.7% of the 60 Ss mastered Generation 2 on passages with organizers while only 38.3% of the 60 mastered Generation 2 on passages without. Of the 54 Ss mastering selection items on passages with organizers, 33.3% also mastered Generation 1 items, while 61.1% were masters on the Generation 2 test. Of the 49 Ss identified as selection masters on passages without organizers, 20.4% and 44.9% also mastered Generation 1 and 2 items respectively. Parallelling the findings in Experiment 1, of the 18 Ss mastering the Generation 1 passages with organizers, 94.4% also mastered the Selection subtest; and of the 10 Ss who mastered the Generation 1 passages without organizers, 100% were masters of the Selection subtest. Of the 34 Generation 2 masters on passages with organizers, and of the 23 Generation 2 masters on passages without organizers, 97.1% and 95.7% were also Selection masters on the corresponding passage types.

Discussion

In both experiments, substantial differences were found between selection and generation response scores. Even when the passages were organized explicitly with topic sentences, the differences remained significant. In general, the Selection and Generation 1 differences were not unexpected, for the criterion that the responses be written as complete sentences may have been unrealistic for a fifth-grade population. The Selection and Generation 2 differences, however, were surprising. Across experiments, less than 50% of the Ss who mastered the selection subtest (that is, responded correctly to 8 of 10 items) also mastered the Generation 2 test. This means that less than half of the Ss who successfully **selected** the correct main idea from among 4 options were also able to **generate** a general response with equal success.

The results of these experiments indicate that selecting and generating passage main idea are different tasks. Even more importantly, they suggest that a test which measures ability to **select** a main idea is probably not measuring the child's ability to **generate** a main idea independently, which is, as stated in the introduction, the major objective of main idea instruction. Interestingly, however, the reverse seems to be true. That is, subjects who can generate a main idea can also select the main idea successfully. The obvious difference between selecting and generating the main idea in this study raises some interesting questions about the appropriateness of a selection test format for determining ability to comprehend main idea independently.

A major weakness of many objective test items is that distractors (i.e., incorrect options) are frequently unfamiliar terms, requiring the child simply to discriminate familiar from unfamiliar words in order to select the correct option. In this study, the incorrect options were terms taken from the passage. Consequently, the child had to discriminate passage details from passage main idea. From the results of these experiments, where the subjects did **more** than simply discriminate familiar from unfamiliar terms, it is obvious that selection test-taking performance is superior to generating performance.

Tests will probably continue to measure the ability to comprehend the main idea via a selection test format, primarily because multiple-choice tests are easier and quicker to score. In the meantime, the classroom teacher might do well to consider augmenting this testing with informal exercises that require the child to generate main idea independently.

References

Farr, R. *Reading: What can be measured?* Newark, Delaware: International Reading Association, 1969.

Osenburg, F.C. Concerning objective reading tests, a minority opinion. *Journal of Developmental Reading,* 1962, 5, 275-279.

Van Blaricom, G. and White, S. The Effect of Passage Organization on Main Idea Comprehension at Three Response Levels. Paper presented at the 1976 International Reading Association Convention in Anaheim.

GENERATION VERSUS SELECTION: SENTENCE MEANING

KARLYN KAMM
University of Wisonsin–Madison

ABSTRACT

The primary focus of this investigation into Sentence Meaning skills was to determine whether the skills of selecting a restatement of a sentence and generating a restatement of a sentence were the same. The nature of the investigation was pragmatic, since it was done in response to several questions formulated during the development of the Comprehension component of the *Wisconsin Design for Reading Skill Development*. Restating a sentence was included in the *Design's* Sentence Meaning strand since it was considered a viable way for a learner to demonstrate his understanding of a complete unit of thought. Change of word order and substitution (use of synonyms) were two ways idientified for a learner to restate a sentence.

A secondary focus of the study was to compare the results with those of a main idea study in which the selection and generation skills were found not to be the same. Correlational analysis was used to test for relationships among the selection and generation tasks. The results indicated that the skills of selecting and generating a restatement of a sentence were not dissimilar. The implication for classroom instruction, however, in light of the findings in the main idea study, seems to be that both selection and generation skills should receive equal emphasis.